

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:04:13 ; Search time 29.1819 Seconds
(without alignments)
1120.478 Million cell updates/sec

Title: US-09-887-855-2_COPY_22_227

Perfect score: 1115

Sequence: 1 ATGRLSGQPCVCRGGTQRC.....EEDAKTKFKESREALNLAY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

| | |
|-----|--|
| 1: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:* |
| 2: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:* |
| 3: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:* |
| 4: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:* |
| 5: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:* |
| 6: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:* |
| 7: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:* |
| 8: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:* |
| 9: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:* |
| 10: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:* |
| 11: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:* |
| 12: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:* |
| 13: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:* |
| 14: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:* |
| 15: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:* |
| 16: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:* |
| 17: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:* |
| 18: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:* |
| 19: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:* |
| 20: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:* |
| 21: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:* |
| 22: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:* |
| 23: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:* |
| 24: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|----|-------------|
| 1 | 1115 | 100.0 | 374 | 21 | AAV93948 |
| 2 | 1115 | 100.0 | 374 | 22 | AAE03651 |
| 3 | 1115 | 100.0 | 374 | 23 | ABG66680 |
| 4 | 1115 | 100.0 | 374 | 23 | ABB90203 |
| 5 | 1115 | 100.0 | 387 | 22 | AAAM25796 |
| 6 | 1105 | 99.1 | 374 | 21 | AAV91490 |
| 7 | 1101 | 98.7 | 382 | 20 | AAV13367 |
| 8 | 1101 | 98.7 | 382 | 22 | AAU29033 |
| 9 | 1101 | 98.7 | 382 | 22 | AAAB0235 |

| | | | | | | |
|----|-------|------|-----|----|----------|--------------------|
| 10 | 1101 | 98.7 | 382 | 24 | ABU69645 | Novel human secret |
| 11 | 1101 | 98.7 | 382 | 24 | ABU71121 | Human PRO234 prote |
| 12 | 1101 | 98.7 | 382 | 24 | ABU71468 | Human PRO polypept |
| 13 | 1101 | 98.7 | 382 | 24 | ABU71914 | Human secreted/tra |
| 14 | 1101 | 98.7 | 382 | 24 | ABU65578 | Human secreted/tra |
| 15 | 1101 | 98.7 | 382 | 24 | ABU65911 | Novel human secret |
| 16 | 1101 | 98.7 | 382 | 24 | ABU67368 | Human secreted pro |
| 17 | 1101 | 98.7 | 382 | 24 | ABU67415 | Human secreted/tra |
| 18 | 1101 | 98.7 | 382 | 24 | ABU64522 | Human secreted/tra |
| 19 | 1101 | 98.7 | 382 | 24 | ABU65273 | Human PRO polypept |
| 20 | 1101 | 98.7 | 382 | 24 | ABU58409 | Human PRO polypept |
| 21 | 1101 | 98.7 | 382 | 24 | ABU55945 | Human secreted/tra |
| 22 | 1101 | 98.7 | 382 | 24 | ABU56940 | Human PRO polypept |
| 23 | 1101 | 98.7 | 382 | 24 | ABU54370 | Human secreted/tra |
| 24 | 1101 | 98.7 | 382 | 24 | ABU10519 | Human secreted/tra |
| 25 | 561.5 | 50.4 | 273 | 21 | AAU18913 | A novel polypeptid |
| 26 | 561.5 | 50.4 | 273 | 22 | AAU12441 | Human PRO1890 poly |
| 27 | 561.5 | 50.4 | 273 | 22 | AAU73309 | Human C-type lecti |
| 28 | 561.5 | 50.4 | 273 | 22 | AAU87609 | Human PRO1890. Ho |
| 29 | 561.5 | 50.4 | 273 | 23 | ABG95934 | Human secreted/tra |
| 30 | 561.5 | 50.4 | 273 | 23 | ABG95554 | Human angiogenesis |
| 31 | 561.5 | 50.4 | 273 | 23 | ABG84948 | Human PRO1890 prot |
| 32 | 561.5 | 50.4 | 273 | 24 | ABU69084 | Human PRO polypept |
| 33 | 561.5 | 50.4 | 273 | 24 | ABU69107 | Human PRO polypept |
| 34 | 561.5 | 50.4 | 273 | 24 | ABU71589 | Human secreted pol |
| 35 | 561.5 | 50.4 | 273 | 24 | ABU72035 | Novel human secret |
| 36 | 561.5 | 50.4 | 273 | 24 | ABU72192 | Human PRO polypept |
| 37 | 561.5 | 50.4 | 273 | 24 | ABU66839 | Human PRO polypept |
| 38 | 561.5 | 50.4 | 273 | 24 | ABU67115 | Human secreted/tra |
| 39 | 561.5 | 50.4 | 273 | 24 | ABU59920 | Novel secreted and |
| 40 | 534.5 | 47.9 | 274 | 23 | ABP69211 | Human polypeptide |
| 41 | 504.5 | 45.2 | 232 | 22 | ABG94192 | Human protein sequ |
| 42 | 275 | 24.7 | 102 | 22 | ABG51278 | Human liver peptid |
| 43 | 275 | 24.7 | 102 | 22 | ABG21788 | Protein #3787 enco |
| 44 | 275 | 24.7 | 102 | 22 | AAU29941 | Peptide #3978 enco |
| 45 | 240 | 21.5 | 81 | 21 | AAV91643 | Human secreted pro |

ALIGNMENTS

RESULT 1
AAV93948
ID AAV93948 standard; Protein; 374 AA.
AC AAV93948;
XX
XX 03-OCT-2000 (first entry)
DT
DE Amino acid sequence of a lectin s83939 polypeptide.
XX
XX Human; lectin s83939; chromosome 11; gene therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT Peptide 1..21 /note= "signal peptide"
FT Domain 22..227 /note= "extracellular coding region"
FT Domain 228..248 /note= "predicted transmembrane domain"
FT Domain 249..374 /note= "predicted cytoplasmic or intracellular domain"

XX
XX WO200039296-A1.
XX
XX 06-JUL-2000.
XX
XX 22-DEC-1999; 99WO-US30523.
XX
XX 23-DEC-1998; 98US-0113820.
XX

PA (IMMV) IMMUNEX CORP.
 XX Anderson DA;
 XX WPI; 2000-452394/39.
 DR N-PSDB; AAA57382.
 XX ss3939 nucleic acids, polypeptides and antibodies, useful for
 PT identifying human chromosome 11 and diseases associated with it -
 XX
 PS Claim 12; Page 8; 73pp; English.
 XX
 CC The present sequence represents a human lectin ss3939 polypeptide. The
 CC polynucleotide sequence is a source of probes, which may be used
 CC to identify nucleic acids encoding ss3939 proteins, to identify human
 CC chromosome number 11, to map genes on human chromosome number 11, to
 CC identify diseases associated with chromosome 11, as single-stranded
 CC sense or antisense oligonucleotides to inhibit expression of
 CC polypeptides encoded by the ss3939 gene, and for gene therapy. The
 CC ss3939 polypeptides may be useful for developing treatments for
 CC diseases (none specified) associated with defective or insufficient
 CC amounts of the polypeptides. The antibodies may be useful for
 CC detecting the presence of ss3939 polypeptides.
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 1115; DB 21; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.5e-105;
 Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGRLSGQVCRGGTQRCYKVIYFHDTSRLNFEAKEACRRDGGQLVSISEDEQKL 60
 Db 22 ATGRLSGQVCRGGTQRCYKVIYFHDTSRLNFEAKEACRRDGGQLVSISEDEQKL 81
 QY 61 IEKFTENILPSGDFWIGLRREKQSNSTACODLYAWTDGSIQPRNRYVDEPSCGSEV 120
 Db 82 IEKFTENILPSGDFWIGLRREKQSNSTACODLYAWTDGSIQPRNRYVDEPSCGSEV 141
 QY 121 CVVMYHQPSAPAGIGPYFQWNNDRCKNNFKICKYDEKPAVPREAGETETLTPV 180
 Db 142 CVVMYHQPSAPAGIGPYFQWNNDRCKNNFKICKYDEKPAVPREAGETETLTPV 201
 QY 181 LPEETOEDAKKTFKESREAAANLAY 206
 Db 202 LPEETOEDAKKTFKESREAAANLAY 227
 RESULT 2
 ID AAE03651 standard; Protein; 374 AA.
 AC AAE03651;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE Human extracellular matrix and cell adhesion molecule-15 (XMAD-15).
 XX
 KW Human; extracellular matrix and cell adhesion molecule; XMAD;
 KW gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
 KW Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
 KW sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma;
 KW inflammatory disorder; acquired immune deficiency syndrome; AIDS;
 KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
 KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
 KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
 KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
 KW infection; cell proliferative disorder; actinic keratosis; myeloma;
 KW arteriosclerosis; neutropenic; anticonvulsant; antithyroid; nephrotropic;
 KW neuroprotective; dermatological.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 Query Match 100.0%; Score 1115; DB 22; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.5e-105;
 Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Peptide 1..24
 Protein 25..374
 /note= "Mature human extracellular matrix and cell
 adhesion molecule (XMAD)"
 Domain 46..63
 /note= "C-type lectin domain"
 Domain 163..176
 /note= "C-type lectin domain"
 Domain 224..247
 /note= "Transmembrane motif"
 Domain 328..348
 /note= "Transmembrane motif"
 WO200142285-A2.
 14-JUN-2001.
 05-DEC-2000; 2000WO-US32990.
 10-DEC-1999; 99US-0172852.
 16-DEC-1999; 99US-0172354.
 (INCY-) INCYTE GENOMICS INC.
 Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
 Baughn MR, Lu DM, Shan P, Au-Young J;
 WPI; 2001-381632/40.
 N-PSDB; AAD08059.
 New human extracellular matrix and cell adhesion molecules and
 polynucleotide sequences encoding them, useful for diagnosis,
 prevention, treatment of genetic, autoimmune and cell proliferative
 disorders -
 Claim 1; Page 108-109; 135pp; English.
 The present sequence is a human extracellular matrix and cell
 adhesion molecule (XMAD). The XMAD is used for screening a compound for
 effectiveness as an agonist or antagonist of XMAD. The identified agonist
 or antagonist are used for treating a disease or condition associated
 with decreased or increased expression of functional XMAD. The
 polynucleotides encoding XMAD are useful in somatic or germline gene
 therapy to correct a genetic deficiency, to express a conditionally
 lethal gene product and to express a protein which affords protection
 against intracellular parasites and also for diagnosis of disorders
 associated with expression of XMAD. They are also used for generating
 hybridisation probes useful in mapping the naturally occurring genomic
 sequences and to create knock in humanised animals (pigs) or transgenic
 animals (mice or rats) to model human diseases. Oligonucleotide or longer
 fragments derived from the polynucleotide sequences may be used as
 elements on a microarray. Antibodies which specifically bind XMAD may be
 used for the diagnosis of disorders associated with the expression of
 XMAD, or in assays to monitor patients being treated with XMAD. Diseases
 diagnosed, prevented or treated include genetic disorders such as
 adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
 disease, myotonic dystrophy, sickle cell anaemia, thalassaemia,
 autoimmune/inflammatory disorders such as acquired immune deficiency
 syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
 atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,
 glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
 osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
 bacterial, fungal, parasitic, protozoal and helminthic infections and
 cell proliferative disorders such as actinic keratosis, arteriosclerosis
 and cancer including breast, bladder, bone marrow, brain and uterus
 cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.
 Sequence 374 AA;

QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLLNFEEAKEACRRDGGQLVSISEDEQKL 60
DB 22 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLLNFEEAKEACRRDGGQLVSISEDEQKL 81
QY 61 IEKFIENLLPSDGGFWIGLRRRREKQSNSTACQDLYAWTDGSIQSFNRYVDEPSCGSEV 120
DB 82 IEKFIENLLPSDGGFWIGLRRRREKQSNSTACQDLYAWTDGSIQSFNRYVDEPSCGSEV 141
QY 121 CVMYHQPAPAGIGGYPYFQWMDRCNMKNFICKYSDEKPAVPSREASGEETELTPV 180
DB 142 CVMYHQPAPAGIGGYPYFQWMDRCNMKNFICKYSDEKPAVPSREASGEETELTPV 201
QY 181 LPETQEDAKKTFKESREAAALNLAY 206
DB 202 LPETQEDAKKTFKESREAAALNLAY 227

RESULT 3
ABG66680
ID ABG66680 standard; Protein: 374 AA.
XX AC ABG66680;
XX DT 30-AUG-2002 (first entry)
XX DE Human novel polypeptide #15.

Human; inflammatory condition; shock; sepsis; immune response;
cancer; wound healing; central nervous system disease; haematopoiesis;
peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
bone degenerative disorder; periodontal disease; reperfusion injury;
lung fibrosis; liver fibrosis; autoimmune disease; bacterial infection;
allergic condition; thrombolytic; thrombosis; coagulation disorder;
fungal infection.

OS Homo sapiens.
XX WO200244340-A2.
XX 06-JUN-2002.
XX 30-NOV-2001; 2001WO-US47004.
XX 30-NOV-2000; 2000US-0028952.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI Yamazaki V, Ujwal ML, Drmanac RT;
XX WPI: 2002-508509/54.
XX N-PSDB; ABK94904.
PT Novel nucleic acids and polypeptides for diagnosis, treatment of
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT disorders, cancer and promoting wound healing -
XX Claim 10; Page 579-580; 672pp; English.

CC The invention relates to human novel polynucleotides and associated
CC polypeptides. The polynucleotides and polypeptides are useful for
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC and cancer and for promoting wound healing. The sequences are used to
CC induce the proliferation of neural cells and regeneration of nerve and
CC brain tissue, and are useful for the treatment of central and peripheral
CC nervous system diseases and neuropathies, such as Alzheimer's disease,
CC Parkinson's disease, Huntington's disease and amyotrophic lateral
CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid

cell disorders and platelet disorders such as thrombocytopenia,
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic
CC and coagulation disorders. Sequences ABG66666-ABG666758 represent human
CC novel polypeptides of the invention.

XX Sequence 374 AA;

Query Match 100.0%; Score 1115; DB 23; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.5e-105;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLLNFEEAKEACRRDGGQLVSISEDEQKL 60
DB 22 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLLNFEEAKEACRRDGGQLVSISEDEQKL 81
QY 61 IEKFIENLLPSDGGFWIGLRRRREKQSNSTACQDLYAWTDGSIQSFNRYVDEPSCGSEV 120
DB 82 IEKFIENLLPSDGGFWIGLRRRREKQSNSTACQDLYAWTDGSIQSFNRYVDEPSCGSEV 141
QY 121 CVMYHQPAPAGIGGYPYFQWMDRCNMKNFICKYSDEKPAVPSREASGEETELTPV 180
DB 142 CVMYHQPAPAGIGGYPYFQWMDRCNMKNFICKYSDEKPAVPSREASGEETELTPV 201
QY 181 LPETQEDAKKTFKESREAAALNLAY 206
DB 202 LPETQEDAKKTFKESREAAALNLAY 227

RESULT 4
ABB90203
ID ABB90203 standard; Protein: 374 AA.
XX AC ABB90203;
XX DT 24-MAY-2002 (first entry)
XX DE Human polypeptide SEQ ID NO 2579.

Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.

OS Homo sapiens.
XX WO200190304-A2.
XX 29-NOV-2001.
XX 18-MAY-2001; 2001WO-US16450.
XX 19-MAY-2000; 2000US-205515P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Birse CE, Rosen CA;
XX WPI: 2002-122018/16.
XX N-PSDB; ABL90612.

Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -

XX PS Claim 11; SEQ ID NO 2579; 2081pp + Sequence Listing; English.

XX CC The invention relates to novel genes (ABL89449-ABL90853) and proteins

XX CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating

XX CC medical conditions e.g. by protein or gene therapy. The genes are

XX CC isolated from a range of human tissues disclosed in the specification.

XX CC The nucleic acids, proteins, antibodies and (ant)agonists are useful

XX CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast

XX CC and ovarian cancer and other cancers of the adrenal gland, bone, bone

XX CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;

XX CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune

XX CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

XX CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative

XX CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;

XX CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and

XX CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

XX CC and parasitic infections.

XX CC Note: The sequence data for this patent did not form part of the

XX CC printed specification, but was obtained in electronic format directly

XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 374 AA;

Query Match 100.0%; Score 1115; DB 23; Length 374;

Best Local Similarity 100.0%; Pred. No. 3.5e-105;

Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSISEDEQKL 60

DB 22 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSISEDEQKL 81

QY 61 IEKFTENLLPSDGFWIGLRREEKQSNSTACODLYAWTDGSIOPRNMVDPSCGSEV 120

DB 82 IEKFTENLLPSDGFWIGLRREEKQSNSTACODLYAWTDGSIOPRNMVDPSCGSEV 141

QY 121 CVVMYHQPAPAGIGGPTMFWQNDRCNMKNFKICKYDEKPAVPSRAEAGSETLTPV 180

DB 142 CVVMYHQPAPAGIGGPTMFWQNDRCNMKNFKICKYDEKPAVPSRAEAGSETLTPV 201

QY 181 LPEETQEDAKKTFKESREAAINLAY 206

DB 202 LPEETQEDAKKTFKESREAAINLAY 227

RESULT 5

ID AAM25796 standard; Protein; 387 AA.

XX AC AAM25796;

XX DT 16-OCT-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:1311.

XX KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;

XX KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;

XX KW antibacterial; endocrine; cardiant; central nervous system; virucide;

XX KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;

XX KW antiagregant; haemostatic; vulnary; antidiabetic; osteopathic; eczema;

XX KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;

XX KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;

XX KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;

XX KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;

XX KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;

XX KW genetic disease; haematopoietic disorder; platelet disorder; asthma;

XX KW thrombocytopenia; osteoporosis; severe combined immunodeficiency;

XX KW allergic rhinitis; diabetes; multiple sclerosis; depression;

XX KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;

XX KW neurological disorder.

OS Homo sapiens.

XX

PN WO200153455-A2.

XX PD 26-JUL-2001.

XX PF 22-DEC-2000; 2000WO-US35017.

XX PR 23-DEC-1999; 99US-0471275.

XX PR 21-JAN-2000; 2000US-048725.

XX PR 25-APR-2000; 2000US-0552317.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-457603/49.

XX DR N-PSDB; AAM99737.

XX DR Isolated human polynucleotides encoding polypeptides, useful for the

XX PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX XX Claim 20; Page 272; 1217pp; English.

XX CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to

XX CC AAM25963. The proteins can have activities based on the tissues and

XX CC cells they are expressed in, such as: antiinflammatory; antirheumatic;

XX CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;

XX CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;

XX CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnary;

XX CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;

XX CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides

XX CC encoding them can be used in gene therapy, antisense therapy and vaccine

XX CC production. The proteins and polynucleotides are useful for screening for

XX CC agonists or antagonists of a protein and for the treatment and diagnosis

XX CC of disorders associated with the activity of a protein e.g. inflammation,

XX CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,

XX CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal

XX CC infections, autoimmunity, genetic diseases, haematopoietic disorders,

XX CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,

XX CC osteoporosis, severe combined immunodeficiency, eczema, allergic

XX CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,

XX CC Alzheimer's disease, Parkinson's disease, neurodegenerative and

XX CC neurological disorders.

XX SQ Sequence 387 AA;

Query Match 100.0%; Score 1115; DB 22; Length 387;

Best Local Similarity 100.0%; Pred. No. 3.6e-105;

Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSISEDEQKL 60

DB 35 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSISEDEQKL 94

QY 61 IEKFTENLLPSDGFWIGLRREEKQSNSTACODLYAWTDGSIOPRNMVDPSCGSEV 120

DB 95 IEKFTENLLPSDGFWIGLRREEKQSNSTACODLYAWTDGSIOPRNMVDPSCGSEV 154

QY 121 CVVMYHQPAPAGIGGPTMFWQNDRCNMKNFKICKYDEKPAVPSRAEAGSETLTPV 180

DB 155 CVVMYHQPAPAGIGGPTMFWQNDRCNMKNFKICKYDEKPAVPSRAEAGSETLTPV 214

QY 181 LPEETQEDAKKTFKESREAAINLAY 206

DB 215 LPEETQEDAKKTFKESREAAINLAY 240

RESULT 6

AA91490

ID AA91490 standard; Protein; 374 AA.

XX AC AA91490;

XX

DT XX 29-JUN-2000 (first entry)
 DE Human secreted protein sequence encoded by gene 40 SEQ ID NO:163.
 XX
 XX Human: secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW antiHIV; antinflammatory; nontropic; neuroprotective; antiallergic;
 KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
 KW antiparatic; cardiant; gene therapy; cancer; neurological disorder;
 KW immune disease; inflammation; blood disorder; tumour.
 XX
 OS Homo sapiens.
 XX
 XX W0200006698-A1.
 PN 10-FEB-2000.
 XX
 XX 29-JUL-1999; 99WO-US17130.
 PF
 XX 30-JUL-1998; 98US-0094657.
 PR 05-AUG-1998; 98US-0095486.
 PR 06-AUG-1998; 98US-0095454.
 PR 06-AUG-1998; 98US-0095455.
 PR 12-AUG-1998; 98US-0096319.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Komatsoulis GA, Rosen SM, Duan R, Moore PA, Shi Y;
 PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
 PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
 DR WPI; 2000-195282/17.
 DR N-PSDB; AAA26385.
 XX
 XX New isolated human genes and the secreted polypeptides they encode,
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders -
 XX
 XX Claim 11; Page 483-484; 634pp; English.
 XX
 XX The polynucleotide sequences given in AAA26346 to AAA26458 encode the
 CC human secreted proteins given in AAY91451 to AAY91691. The human secreted
 CC proteins can have activities based on the tissues and cells they are
 CC expressed in. Examples of the activities are: cytostatic;
 CC immunosuppressive; antiHIV; antinflammatory; nontropic; neuroprotective;
 CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
 CC antiasthma; antiparatic; and cardiant. The polynucleotides and their
 CC corresponding secreted proteins are useful for preventing, treating or
 CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
 CC pathological conditions can be diagnosed by determining the amount of the
 CC proteins in a sample or by determining the presence of mutations in the
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 CC cardiovascular disorders, reproductive disorders, gastrointestinal
 CC disorders, respiratory disorders and metabolic disorders. The proteins
 CC or polynucleotides can also be used as food additives or preservatives.
 CC The proteins are also useful for identifying their binding partners.
 CC AAA26337 to AAA26345 and AAY91450 are sequences used in the
 CC exemplification of the present invention.
 XX
 XX Sequence 374 AA;
 SQ
 Query Match 99.1%; Score 1105; DB 21; Length 374;
 Best Local Similarity 99.0%; Pred. No. 3.7e-104;
 Matches 204; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATGRLSSQPVCRGQTQPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSISEDEQKL 60

Db 22 ATGRLSSQPVCRGQTQPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSISEDEQKL 81
 Qy 61 IEKFIENLLPSDGDGFWIGLRREBKQSNSTACQDLIYAWTDGSIQFRNRYVDEPSCGSEV 120
 Db 82 IEXFIENLLPSDGDGFWIGLRREBKQSNSTXCDLIYAWTDGSIQFRNRYVDEPSCGSEV 141
 Qy 121 CVVMYHOPSPAGIGGPPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPV 180
 Db 142 CVVMYHOPSPAGIGGPPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPV 201
 Qy 181 LPETQEDAKKTFKESREAAALNLAY 206
 Db 202 LPETQEDAKKTFKESREAAALNLAY 227
 RESULT 7
 AAY13367
 ID AAY13367 standard; Protein; 382 AA.
 XX
 AC AAY13367;
 XX
 DT 25-JUN-1999 (first entry)
 DE
 DE Amino acid sequence of protein PRO234.
 XX
 KW Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW congenital microvillus atrophy; skin disease; cell growth;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
 KW anti-thrombotic; wound healing; tissue repair.
 XX
 OS Homo sapiens.
 XX
 PN W09914328-A2.
 XX
 PD 25-MAR-1999.
 XX
 PF 16-SEP-1998; 98WO-US19330.
 XX
 XX 25-NOV-1997; 97US-0066840.
 PR 17-SEP-1997; 97US-0059113.
 PR 17-SEP-1997; 97US-0059115.
 PR 17-SEP-1997; 97US-0059117.
 PR 17-SEP-1997; 97US-0059119.
 PR 17-SEP-1997; 97US-0059121.
 PR 17-SEP-1997; 97US-0059122.
 PR 17-SEP-1997; 97US-0059184.
 PR 18-SEP-1997; 97US-0059263.
 PR 18-SEP-1997; 97US-0059266.
 PR 15-OCT-1997; 97US-0062125.
 PR 17-OCT-1997; 97US-0062285.
 PR 17-OCT-1997; 97US-0062287.
 PR 21-OCT-1997; 97US-0063486.
 PR 24-OCT-1997; 97US-0062814.
 PR 24-OCT-1997; 97US-0062816.
 PR 24-OCT-1997; 97US-0063045.
 PR 24-OCT-1997; 97US-0063120.
 PR 24-OCT-1997; 97US-0063121.
 PR 24-OCT-1997; 97US-0063127.
 PR 24-OCT-1997; 97US-0063128.
 PR 27-OCT-1997; 97US-0063329.
 PR 27-OCT-1997; 97US-0063327.
 PR 28-OCT-1997; 97US-0063541.
 PR 28-OCT-1997; 97US-0063542.
 PR 28-OCT-1997; 97US-0063544.
 PR 28-OCT-1997; 97US-0063549.
 PR 28-OCT-1997; 97US-0063550.
 PR 28-OCT-1997; 97US-0063564.
 PR 29-OCT-1997; 97US-0063435.
 PR 29-OCT-1997; 97US-0063704.

PR 29-OCT-1997; 97US-0063732.
 PR 29-OCT-1997; 97US-0063738.
 PR 29-OCT-1997; 97US-0063734.
 PR 29-OCT-1997; 97US-0064215.
 PR 29-OCT-1997; 97US-0063735.
 PR 31-OCT-1997; 97US-0063870.
 PR 31-OCT-1997; 97US-0064103.
 PR 03-NOV-1997; 97US-0064248.
 PR 07-NOV-1997; 97US-0064809.
 PR 12-NOV-1997; 97US-0065186.
 PR 17-NOV-1997; 97US-0065846.
 PR 18-NOV-1997; 97US-0065693.
 PR 21-NOV-1997; 97US-0066120.
 PR 21-NOV-1997; 97US-0066364.
 PR 24-NOV-1997; 97US-0066772.
 PR 24-NOV-1997; 97US-0066466.
 PR 24-NOV-1997; 97US-0066770.
 PR 24-NOV-1997; 97US-0066511.
 PR 24-NOV-1997; 97US-0066453.
 XX (GETH) GENENTECH INC.
 PA
 XX
 PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 XX WPI; 1999-229533/19.
 DR N-PSDB; AAX52238.
 DR
 XX
 PT New isolated human genes and polypeptides used in, e.g. treatment of
 PT gastrointestinal ulceration
 XX
 PS Claim 12; Fig 50; 320pp; English.
 XX
 CC AAY13344-403 represent secreted and transmembrane human proteins.
 CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to
 CC growth or survival of nerve cells including Parkinson's disease,
 CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
 CC for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO333 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair; PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.
 XX
 SQ Sequence 382 AA;
 Query Match 98.7%; Score 1101; DB 20; Length 382;
 Best Local Similarity 96.3%; Pred. No. 9.6e-104;
 Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 ATGRLLS-----GQPVCRGTQPCVKVIYFHDTSRLNFEEAKEACRRDGGQLVSI 52
 DB 22 ATGRLLSASDLRLGQGPVCRGTQPCVKVIYFHDTSRLNFEEAKEACRRDGGQLVSI 81
 QY 53 ESEDEKLEKLEIENLLPSDGFNGLRRREKQSNSTACQDIYAWTDGSIQFRNYYVD 112
 DB 82 ESEDEKLEKLEIENLLPSDGFNGLRRREKQSNSTACQDIYAWTDGSIQFRNYYVD 141
 QY 113 EPSCSCEVCVMYHOPSPAGIGGPMYFQWDDRCNMKNFKICKYSDEKPAVPSREAEGE 172
 DB 142 EPSCSCEVCVMYHOPSPAGIGGPMYFQWDDRCNMKNFKICKYSDEKPAVPSREAEGE 201
 QY 173 ETELTTPVLPEETQBEDAKKTPKESREAAALNLAY 206

Db 202 ETELTTPVLPEETQBEDAKKTPKESREAAALNLAY 235

RESULT 8

AAU29033
ID AAU29033 standard; Protein; 382 AA.

XX AAU29033;

XX 18-DEC-2001 (first entry)

XX Human PRO polypeptide sequence #10.

XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200168848-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06520.
 XX
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 03-MAR-2000; 2000US-187202P.
 PR 06-MAR-2000; 2000US-186968P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 14-MAR-2000; 2000US-189328P.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 21-MAR-2000; 2000US-190828P.
 PR 21-MAR-2000; 2000US-191007P.
 PR 21-MAR-2000; 2000US-191048P.
 PR 21-MAR-2000; 2000US-191314P.
 PR 28-MAR-2000; 2000US-192655P.
 PR 29-MAR-2000; 2000US-193032P.
 PR 30-MAR-2000; 2000US-193053P.
 PR 04-APR-2000; 2000WO-US08439.
 PR 04-APR-2000; 2000US-194449P.
 PR 11-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 03-MAY-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 22-MAY-2000; 2000WO-US13705.
 PR 30-MAY-2000; 2000WO-US14042.
 PR 02-JUN-2000; 2000WO-US14941.
 PR 05-JUN-2000; 2000WO-US15264.
 PR 28-JUL-2000; 2000US-209832P.
 PR 22-AUG-2000; 2000US-064484P.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX WPI; 2001-602746/68.

DR N-PSDB; AAS45934.

XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the

PT presence of tumours, such as prostate and breast tumours, in mammals and

PT to screen for modulators of the compounds -

XX Claim 11; Fig 20; 774pp; English.

XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.

CC The PRO polypeptides and their associated nucleic acids can be used to

CC detect the presence of a tumour in a mammal by comparing the level of

CC expression of a PRO polypeptide in a test sample of cells from the animal

CC and a control sample of normal cells, whereby a higher level of

CC expression in the test sample indicates the presence of a tumour in the

CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats

CC and rabbits but are preferably human. The polypeptides can be used to

CC stimulate tumour necrosis factor (TNF) alpha release from human blood,

CC when contacted with it. A specific polypeptide can be used to stimulate

CC the proliferation or differentiation of chondrocyte cells. The PRO

CC proteins can be used to determine the presence of tumours and also

CC susceptibility to tumour development, particularly adrenal, lung, colon,

CC breast, prostate, rectal, cervical, or liver tumours, in mammalian

CC subjects. The oligonucleotide probes specific for the PRO nucleic acids

CC can be used for genetic analysis of individuals with genetic disorders.

XX SQ Sequence 382 AA;

Query Match 98.7%; Score 1101; DB 22; Length 382;

Best Local Similarity 96.3%; Pred. No. 9.6e-104;

Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSI 52

DB 22 ATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSI 81

QY 53 ESEDEQKLIKFIENLLPSDGFWIGLRREEKQSNSTACQDLYAWTDGSIQFRNMYVD 112

DB 82 ESEDEQKLIKFIENLLPSDGFWIGLRREEKQSNSTACQDLYAWTDGSIQFRNMYVD 141

QY 113 EPSCGSEVCVMYHQPAPAGIGPYMFQWDDRCNKNPFICKYSDKPAVPSREAGE 172

DB 142 EPSCGSEVCVMYHQPAPAGIGPYMFQWDDRCNKNPFICKYSDKPAVPSREAGE 201

QY 173 ETELTTPVLPEETQEDAKTKFKESREAAALNLAY 206

DB 202 ETELTTPVLPEETQEDAKTKFKESREAAALNLAY 235

RESULT 9

AA880235

ID AAB80235 standard; Protein; 382 AA.

XX AAB80235;

DT 24-APR-2001 (first entry)

XX Human PRO234 protein.

XX Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;

KW antiparkinsonian nootropic; neuroprotective; vulnerary; cardiant;

KW antiangiogenic; vasotropic; antiasthmatic; antithematic; cancer;

KW antiarthritic; antinfertility; antidiabetic; antiviral; diabetes;

KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;

KW ischaemia; inflammation.

XX Homo sapiens.

XX WO200104311-A1.

XX 18-JAN-2001.

XX 22-FEB-2000; 2000WO-US04414.

XX

PR 07-JUL-1999; 99US-0143048.

PR 26-JUL-1999; 99US-0145698.

PR 28-JUL-1999; 99US-0146222.

PR 13-SEP-1999; 99WO-US20594.

PR 15-SEP-1999; 99WO-US21094.

PR 15-SEP-1999; 99WO-US21547.

PR 05-OCT-1999; 99WO-US23089.

PR 29-NOV-1999; 99WO-US28214.

PR 30-NOV-1999; 99WO-US28313.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 20-DEC-1999; 99WO-US30999.

PR 05-JAN-2000; 99WO-US00219.

XX (GETH) GENENTECH INC.

PA Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;

PI Fikvaroff E, Fong S, Gao W, Gerber H, Gertitsen ME, Goddard A;

PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin IJ;

PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;

PI Williams PM, Wood WI;

XX WPI; 2001-081051/09.

DR N-PSDB; AAF72396.

XX

PT Sixty one nucleic acids encoding PRO polypeptides which are useful in

PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung

PT squamous cell carcinoma) and neurodegenerative diseases (e.g.

PT Alzheimer's disease) -

XX Claim 1; Fig 50; 393pp; English.

XX

CC The present sequence is one of sixty one novel secreted and

CC transmembrane PRO polypeptides. The PRO polypeptides are

CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung

CC squamous cell carcinoma), gastrointestinal disorders (e.g.

CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,

CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.

CC endometrial bleeding angiogenesis, ischaemia such as coronary

CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,

CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and

CC diabetes and retinal disorders such as retinitis pigmentosum.

CC The PRO nucleic acids have applications in molecular biology, including

CC use as hybridization probes, and in chromosome and gene mapping.

XX SQ Sequence 382 AA;

Query Match 98.7%; Score 1101; DB 22; Length 382;

Best Local Similarity 96.3%; Pred. No. 9.6e-104;

Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSI 52

DB 22 ATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSI 81

QY 53 ESEDEQKLIKFIENLLPSDGFWIGLRREEKQSNSTACQDLYAWTDGSIQFRNMYVD 112

DB 82 ESEDEQKLIKFIENLLPSDGFWIGLRREEKQSNSTACQDLYAWTDGSIQFRNMYVD 141

QY 113 EPSCGSEVCVMYHQPAPAGIGPYMFQWDDRCNKNPFICKYSDKPAVPSREAGE 172

DB 142 EPSCGSEVCVMYHQPAPAGIGPYMFQWDDRCNKNPFICKYSDKPAVPSREAGE 201

QY 173 ETELTTPVLPEETQEDAKTKFKESREAAALNLAY 206

DB 202 ETELTTPVLPEETQEDAKTKFKESREAAALNLAY 235

RESULT 10

ABU69645

ID ABU69645 standard; Protein; 382 AA.

XX

| | | | | | | | | | | | |
|-----------|--|--|--------------|-----------|------------|-----|--------|------|------|----|--|
| CC | protein electrophoresis purposes. The anti-PRO antibodies may be used | | | | | | | | | | |
| CC | in diagnostic assays for PRO, or for the affinity purification of PRO | | | | | | | | | | |
| CC | from recombinant cell culture or natural sources. This is the amino | | | | | | | | | | |
| CC | acid sequence of a novel human PRO polypeptide. | | | | | | | | | | |
| XX | | | | | | | | | | | |
| SQ | Sequence | 382 AA; | | | | | | | | | |
| | Query Match | 98.7%; | Score | 1101; | DB | 24; | Length | 382; | | | |
| | Best Local Similarity | 96.3%; | Pred. No. | 9.6e-104; | | | | | | | |
| | Matches | 206; | Conservative | 0; | Mismatches | 0; | Indels | 8; | Gaps | 1; | |
| Oy | 1 | ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEAEACRRDGGQLVSI | 52 | | | | | | | | |
| Db | 22 | ATGRLLSASDLRGGQVCRGGTQPCYKVIYFHDTSRRLNFEAEACRRDGGQLVSI | 81 | | | | | | | | |
| Oy | 53 | ESEDEQKLIKFIENLLPSDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVD | 112 | | | | | | | | |
| Db | 82 | ESEDEQKLIKFIENLLPSDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVD | 141 | | | | | | | | |
| Oy | 113 | EPSCGSEVCVMYHOPSAAGIGGYPMFQWMDRCNKNMFICKYSDKPAVPSREAEGE | 172 | | | | | | | | |
| Db | 142 | EPSCGSEVCVMYHOPSAAGIGGYPMFQWMDRCNKNMFICKYSDKPAVPSREAEGE | 201 | | | | | | | | |
| Oy | 173 | ETELTTPVLPEETQEDAKKTFKESREAAALNLAY | 206 | | | | | | | | |
| Db | 202 | ETELTTPVLPEETQEDAKKTFKESREAAALNLAY | 235 | | | | | | | | |
| RESULT 11 | | | | | | | | | | | |
| | ABU71121 | | | | | | | | | | |
| ID | ABU71121 | standard; Protein; 382 AA. | | | | | | | | | |
| XX | ABU71121; | | | | | | | | | | |
| DT | 10-JUN-2003 | (first entry) | | | | | | | | | |
| XX | Human | PRO234 protein. | | | | | | | | | |
| KW | Human; PRO; secreted; transmembrane; cytostatic; TNF-alpha; blood; | | | | | | | | | | |
| KW | tumour necrosis factor alpha release; chondrocyte cell; proliferation; | | | | | | | | | | |
| KW | differentiation; tumour; gene therapy. | | | | | | | | | | |
| XX | Homo sapiens. | | | | | | | | | | |
| OS | US2003036143-A1. | | | | | | | | | | |
| XX | 20-FEB-2003. | | | | | | | | | | |
| XX | 02-JUL-2002; 2002US-0187600. | | | | | | | | | | |
| PR | 16-SEP-1998; | 98WO-US19330. | | | | | | | | | |
| PR | 07-OCT-1998; | 98WO-US21141. | | | | | | | | | |
| PR | 01-DEC-1998; | 98WO-US25108. | | | | | | | | | |
| PR | 08-MAR-1999; | 99WO-US05028. | | | | | | | | | |
| PR | 14-MAY-1999; | 99WO-US10733. | | | | | | | | | |
| PR | 02-JUN-1999; | 99WO-US12252. | | | | | | | | | |
| PR | 01-SEP-1999; | 99WO-US20111. | | | | | | | | | |
| PR | 15-SEP-1999; | 99WO-US21090. | | | | | | | | | |
| PR | 01-DEC-1999; | 99WO-US28301. | | | | | | | | | |
| PR | 02-DEC-1999; | 99WO-US28551. | | | | | | | | | |
| PR | 30-DEC-1999; | 99WO-US31274. | | | | | | | | | |
| PR | 05-JAN-2000; | 2000WO-US00219. | | | | | | | | | |
| PR | 18-FEB-2000; | 2000WO-US04341. | | | | | | | | | |
| PR | 18-FEB-2000; | 2000WO-US04342. | | | | | | | | | |
| PR | 22-FEB-2000; | 2000WO-US04414. | | | | | | | | | |
| PR | 24-FEB-2000; | 2000WO-US05004. | | | | | | | | | |
| PR | 01-MAR-2000; | 2000WO-US05601. | | | | | | | | | |
| PR | 02-MAR-2000; | 2000WO-US05841. | | | | | | | | | |
| PR | 15-MAR-2000; | 2000WO-US06884. | | | | | | | | | |
| PR | 30-MAR-2000; | 2000WO-US08439. | | | | | | | | | |
| PR | 17-MAY-2000; | 2000WO-US13705. | | | | | | | | | |
| PR | 22-MAY-2000; | 2000WO-US14042. | | | | | | | | | |
| PR | 30-MAY-2000; | 2000WO-US14941. | | | | | | | | | |

| | | |
|----|--------------|-----------------|
| PR | 02-JUN-2000; | 2000WO-US15264. |
| PR | 28-JUL-2000; | 2000WO-US20710. |
| PR | 24-AUG-2000; | 2000WO-US23328. |
| PR | 08-NOV-2000; | 2000WO-US30952. |
| PR | 01-DEC-2000; | 2000WO-US32678. |
| PR | 20-DEC-2000; | 2000WO-US34956. |
| PR | 28-FEB-2001; | 2001WO-US06520. |
| PR | 01-JUN-2001; | 2001WO-US17800. |
| PR | 20-JUN-2001; | 2001WO-US19692. |
| PR | 29-JUN-2001; | 2001WO-US21066. |
| PR | 09-JUL-2001; | 2001WO-US21735. |
| PR | 29-AUG-2001; | 2001WO-US27099. |
| PR | 18-SEP-1997; | 97US-059263P. |
| PR | 18-SEP-1997; | 97US-059266P. |
| PR | 17-OCT-1997; | 97US-062250P. |
| PR | 21-OCT-1997; | 97US-063486P. |
| PR | 24-OCT-1997; | 97US-063120P. |
| PR | 24-OCT-1997; | 97US-063121P. |
| PR | 28-OCT-1997; | 97US-063540P. |
| PR | 28-OCT-1997; | 97US-063541P. |
| PR | 28-OCT-1997; | 97US-063544P. |
| PR | 28-OCT-1997; | 97US-063564P. |
| PR | 29-OCT-1997; | 97US-063734P. |
| PR | 31-OCT-1997; | 97US-063870P. |
| PR | 31-OCT-1997; | 97US-064103P. |
| PR | 13-NOV-1997; | 97US-065311P. |
| PR | 21-NOV-1997; | 97US-066120P. |
| PR | 24-NOV-1997; | 97US-066466P. |
| PR | 24-NOV-1997; | 97US-066772P. |
| PR | 11-DEC-1997; | 97US-069335P. |
| PR | 12-DEC-1997; | 97US-069425P. |
| PR | 17-DEC-1997; | 97US-069870P. |
| PR | 18-DEC-1997; | 97US-068017P. |
| PR | 10-MAR-1998; | 98US-077450P. |
| PR | 11-MAR-1998; | 98US-077632P. |
| PR | 11-MAR-1998; | 98US-077649P. |
| PR | 20-MAR-1998; | 98US-078886P. |
| PR | 20-MAR-1998; | 98US-078939P. |
| PR | 27-MAR-1998; | 98US-079664P. |
| PR | 27-MAR-1998; | 98US-079786P. |
| PR | 31-MAR-1998; | 98US-080107P. |
| PR | 01-APR-1998; | 98US-080327P. |
| PR | 01-APR-1998; | 98US-080333P. |
| PR | 08-APR-1998; | 98US-081049P. |
| PR | 08-APR-1998; | 98US-081070P. |
| PR | 09-APR-1998; | 98US-081195P. |
| PR | 15-APR-1998; | 98US-081838P. |
| PR | 21-APR-1998; | 98US-082568P. |
| PR | 21-APR-1998; | 98US-082569P. |
| PR | 22-APR-1998; | 98US-082704P. |
| PR | 22-APR-1998; | 98US-082797P. |
| PR | 28-APR-1998; | 98US-083322P. |
| PR | 29-APR-1998; | 98US-083495P. |
| PR | 29-APR-1998; | 98US-083496P. |
| PR | 29-APR-1998; | 98US-083499P. |
| PR | 05-MAY-1998; | 98US-083559P. |
| PR | 05-MAY-1998; | 98US-083666P. |
| PR | 06-MAY-1998; | 98US-084414P. |
| PR | 07-MAY-1998; | 98US-084633P. |
| PR | 07-MAY-1998; | 98US-084640P. |
| PR | 13-MAY-1998; | 98US-084643P. |
| PR | 15-MAY-1998; | 98US-085579P. |
| PR | 15-MAY-1998; | 98US-085580P. |
| PR | 15-MAY-1998; | 98US-085802P. |
| PR | 15-MAY-1998; | 98US-085804P. |
| PR | 18-MAY-1998; | 98US-086023P. |
| PR | 22-MAY-1998; | 98US-086392P. |
| PR | 22-MAY-1998; | 98US-086466P. |
| PR | 28-MAY-1998; | 98US-087088P. |
| PR | 28-MAY-1998; | 98US-087089P. |
| PR | 02-JUN-1999; | 98US-087603P. |
| PR | 02-JUN-1999; | 98US-087753P. |

```
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088036P.
PR 04-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-088909P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091135P.
PR 01-JUL-1998; 98US-091154P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.

PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.
PR 01-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.
PR 10-SEP-1998; 98US-099754P.
PR 10-SEP-1998; 98US-099763P.
PR 10-SEP-1998; 98US-099812P.

Query Match 98.7%; Score 1101; DB 24; Length 382;
Best Local Similarity 96.3%; Pred. No. 9.6e-104;
Matches 206; Conservative 0; Mismatch 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQPVCRGGTQPCYKVIFYFHDTSRRLNPFEEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLRLGGQPVCRGGTQPCYKVIFYFHDTSRRLNPFEEAKEACRRDGGQLVSI 81
QY 53 ESEDEQKLIKEFIENLLPSDGFWIGLRRREKQSNSTACQDLAYMTDGSISQFRNYYVD 112
Db 82 ESEDEQKLIKEFIENLLPSDGFWIGLRRREKQSNSTACQDLAYMTDGSISQFRNYYVD 141
QY 113 BPSGSEVCVVMYHOPSPAGIGGPPYFQWDDRCNMKNFICKYSDEKPAVPSREAGE 172
Db 142 BPSGSEVCVVMYHOPSPAGIGGPPYFQWDDRCNMKNFICKYSDEKPAVPSREAGE 201
QY 173 ETELTPVLPEETQEDAKKTFKESREALNLAY 206
Db 202 ETELTPVLPEETQEDAKKTFKESREALNLAY 235

RESULT 12
ABU71468
ID ABU71468 standard; Protein; 382 AA.
XX
AC ABU71468;
XX
DT 10-JUN-2003 (first entry)
XX
DE Human PRO polypeptide #24.
XX
KW Human; secreted and transmembrane protein; PRO polypeptide; cancer;
KW Alzheimer's disease; ischaemia; cytostatic; neurotropic; vasotropic;
KW neuroprotective.
XX
OS Homo sapiens.
XX
PN US2002192659-A1.
XX
PD 19-DEC-2002.
XX
PF 10-JUL-2001; 2001US-0902853.
XX
PR 10-SEP-1998; 98WO-US18824.
PR 14-SEP-1998; 98WO-US19177.
PR 16-SEP-1998; 98WO-US19330.
PR 17-SEP-1998; 98WO-US19437.
PR 01-DEC-1998; 98WO-US25108.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 05-JAN-2000; 2000WO-US00219.
PR 11-FEB-2000; 2000WO-US03565.
```

```

PR 22-FEB-2000; 2000WO-US04414.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 17-SEP-1997; 97US-059113P.
PR 17-SEP-1997; 97US-059115P.
PR 17-SEP-1997; 97US-059117P.
PR 18-SEP-1997; 97US-059266P.
PR 15-OCT-1997; 97US-062125P.
PR 17-OCT-1997; 97US-062285P.
PR 17-OCT-1997; 97US-062287P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-062814P.
PR 24-OCT-1997; 97US-062816P.
XX PA (GETH ) GENENTECH INC.
XX XX
XX PI Ashkenazi A, Botstein D, Deanovers L, Eaton DL, Ferrara N;
PI Filvaroff E, Fong S, Gao W, Garber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kljavin IJ;
PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
PI Williams PM, Wood WI;
XX XX
XX DR WPI; 2003-361832/34.
XX DR N-PSDB; ACAS8405.
XX XX
XX PT New isolated nucleic acid encoding a PRO polypeptide, e.g. PRO245 or
XX PT PRO1868, useful in molecular biology, chromosome and gene mapping, in
XX PT generating antisense RNA and DNA, and in gene therapy -
XX XX
XX PS Claim 12; Fig 50; 474pp; English.
XX XX
XX CC The present invention relates to the isolation of novel human secreted
XX CC and transmembrane proteins (PRO polypeptides), and the polynucleotide
XX CC sequences encoding them. The polynucleotide sequences are useful in
XX CC molecular biology, as hybridisation probes, in chromosome and gene
XX CC mapping, in generating antisense RNA and DNA, and in gene therapy. The
XX CC polynucleotide sequences may also be used in preparing PRO polypeptides
XX CC by recombinant techniques, and in generating either transgenic animals
XX CC or knock-out animals which, in turn, are useful in the development and
XX CC screening of therapeutically useful reagents. The PRO polypeptides or
XX CC their antibodies are useful in preparing a medicament for treating a
XX CC condition responsive to the polypeptide or antibody, such as cancer,
XX CC Alzheimer's disease or ischaemia, and in various diagnostic assays.
XX CC ABU71445-ABU71505 represent human PRO polypeptides of the invention.
XX XX
XX SQ Sequence 382 AA;
Query Match 98.7%; Score 1101; DB 24; Length 382;
Best Local Similarity 96.3%; Pred. No. 9.6e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
OY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSI 52
Db 22 ATGRLLSASDLRGGQPCVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSI 81
OY 53 ESEDEQKLEKFTENLLPSDGFHWGLRREKQSNSTACQDLYAMTDGSIQFRNYYVD 112
Db 82 ESEDEQKLEKFTENLLPSDGFHWGLRREKQSNSTACQDLYAMTDGSIQFRNYYVD 141
OY 113 EPSCGSEVCVMVTHQSPAGIGYPMFQNDRCNMKNFICKYSDKPAVPSRAEGE 172
Db 142 EPSCGSEVCVMVTHQSPAGIGYPMFQNDRCNMKNFICKYSDKPAVPSRAEGE 201
OY 173 ETELTTPVLPEETQEDAKKTFKESREALNLAY 206
Db 202 ETELTTPVLPEETQEDAKKTFKESREALNLAY 235
RESULT 13
ABU71914
ID ABU71914 standard; Protein; 382 AA.
XX XX
AC ABU71914;

```

```

XX DT 12-JUN-2003 (first entry)
XX DE Human secreted/transmembrane protein PRO234.
XX KW Human; secreted protein; transmembrane protein; PRO;
XX KW gene therapy; chromosome identification; chromosome marker.
XX OS Homo sapiens.
XX PN US2003003530-A1.
XX PD 02-JAN-2003.
XX PF 11-JUL-2001; 2001US-0904011.
XX PR 10-SEP-1998; 98WO-US18824.
XX PR 14-SEP-1998; 98WO-US19177.
XX PR 16-SEP-1998; 98WO-US19330.
XX PR 17-SEP-1998; 98WO-US19437.
XX PR 01-DEC-1998; 98WO-US25108.
XX PR 08-SEP-1999; 99WO-US20594.
XX PR 13-SEP-1999; 99WO-US20944.
XX PR 15-SEP-1999; 99WO-US21090.
XX PR 15-SEP-1999; 99WO-US21547.
XX PR 05-OCT-1999; 99WO-US23089.
XX PR 29-NOV-1999; 99WO-US28214.
XX PR 30-NOV-1999; 99WO-US28313.
XX PR 01-DEC-1999; 99WO-US28301.
XX PR 02-DEC-1999; 99WO-US28564.
XX PR 02-DEC-1999; 99WO-US28565.
XX PR 16-DEC-1999; 99WO-US30095.
XX PR 20-DEC-1999; 99WO-US30911.
XX PR 20-DEC-1999; 99WO-US30999.
XX PR 05-JAN-2000; 2000WO-US00219.
XX PR 11-FEB-2000; 2000WO-US03565.
XX PR 22-FEB-2000; 2000WO-US04414.
XX PR 02-MAR-2000; 2000WO-US05004.
XX PR 02-MAR-2000; 2000WO-US05841.
XX PR 20-MAR-2000; 2000WO-US07377.
XX PR 30-MAR-2000; 2000WO-US08439.
XX PR 02-MAY-2000; 2000WO-US14042.
XX PR 02-JUN-2000; 2000WO-US15264.
XX PR 28-JUL-2000; 2000WO-US20710.
XX PR 24-AUG-2000; 2000WO-US23328.
XX PR 17-SEP-1997; 97US-059113P.
XX PR 17-SEP-1997; 97US-059115P.
XX PR 17-SEP-1997; 97US-059117P.
XX PR 17-SEP-1997; 97US-059119P.
XX PR 17-SEP-1997; 97US-059121P.
XX PR 17-SEP-1997; 97US-059122P.
XX PR 18-SEP-1997; 97US-059184P.
XX PR 18-SEP-1997; 97US-059263P.
XX PR 18-SEP-1997; 97US-059266P.
XX PR 15-OCT-1997; 97US-062125P.
XX PR 17-OCT-1997; 97US-062285P.
XX PR 17-OCT-1997; 97US-062287P.
XX PR 21-OCT-1997; 97US-063486P.
XX PR 24-OCT-1997; 97US-062814P.
XX PR 24-OCT-1997; 97US-062816P.
XX PR 24-OCT-1997; 97US-063045P.
XX PR 24-OCT-1997; 97US-063120P.
XX PR 24-OCT-1997; 97US-063121P.
XX PR 24-OCT-1997; 97US-063127P.
XX PR 24-OCT-1997; 97US-063128P.
XX PR 27-OCT-1997; 97US-063327P.
XX PR 27-OCT-1997; 97US-063329P.
XX PR 28-OCT-1997; 97US-063541P.
XX PR 28-OCT-1997; 97US-063542P.
XX PR 28-OCT-1997; 97US-063544P.
XX PR 28-OCT-1997; 97US-063549P.
XX PR 28-OCT-1997; 97US-063550P.
XX PR 28-OCT-1997; 97US-063564P.

```

PR 29-OCT-1997; 97US-0634335P.
 PR 29-OCT-1997; 97US-063704P.
 PR 29-OCT-1997; 97US-063732P.
 PR 29-OCT-1997; 97US-063734P.
 PR 29-OCT-1997; 97US-063735P.
 PR 29-OCT-1997; 97US-063738P.
 PR 29-OCT-1997; 97US-064215P.
 PR 31-OCT-1997; 97US-063870P.
 PR 31-OCT-1997; 97US-064103P.
 PR 03-NOV-1997; 97US-064248P.
 PR 07-NOV-1997; 97US-064809P.
 PR 12-NOV-1997; 97US-065186P.
 PR 17-NOV-1997; 97US-065846P.
 PR 18-NOV-1997; 97US-065693P.
 PR 21-NOV-1997; 97US-066120P.
 PR 21-NOV-1997; 97US-066364P.
 PR 24-NOV-1997; 97US-066453P.
 PR 24-NOV-1997; 97US-066466P.
 PR 24-NOV-1997; 97US-066511P.
 PR 24-NOV-1997; 97US-066770P.
 PR 24-NOV-1997; 97US-066772P.
 PR 18-SEP-2000; 2000US-0665350.

XX (GETH) GENENTECH INC.

PI Ashkenazi A, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
 PI Filvaroff E, Fong S, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kljavin LJ;
 PI Mather JP, Pan J, Paoi NF, Roy MA, Stewart TA, Tumas D;
 PI Williams PM, Wood WI;
 XX WPI; 2003-329602/31.
 DR N-PSDB; ACA60112.

XX New transmembrane polypeptides and nucleic acids encoding the
 PT polypeptides, useful in gene therapy, in chromosome identification, as
 PT chromosome markers, in generating probes and in tissue typing -

XX Claim 12; Fig 50; 484pp; English.

XX The invention relates to an isolated nucleic acid with at least 80%
 CC nucleic acid sequence identity to a nucleotide sequence encoding one of
 CC 61 secreted/transmembrane polypeptides, or PRO polypeptides or encoding a
 CC PRO protein extracellular domain. Also included are a vector comprising
 CC the PRO nucleic acid, a host cell comprising the vector, producing a PRO
 CC polypeptide (by culturing the host cell for the expression of the PRO
 CC polypeptide, and recovering the PRO polypeptide from the cell culture),
 CC an isolated PRO polypeptide (having at least 80% sequence identity
 CC to: (a) an amino acid sequence selected from the 61 PRO proteins;
 CC (b) an amino acid sequence encoded by a nucleic acid molecule deposited
 CC with an ATCC number (detailed in the specification); or (c) an
 CC extracellular domain of a PRO polypeptide or to a PRO polypeptide lacking
 CC its associated signal peptide), a chimeric molecule comprising a PRO
 CC polypeptide of fused to a heterologous amino acid sequence, an anti-PRO
 CC antibody, detecting a PRO245 or PRO1868 in a sample suspected of
 CC containing the polypeptide, linking a bioactive molecule to a cell
 CC expressing a PRO245 or PRO1868 and modulating at least one biological
 CC activity of a cell expressing a PRO245 or PRO1868. Nucleic acids which
 CC encode PRO can be used to generate either transgenic animals or knock-out
 CC animals which may be used in the development and screening of
 CC therapeutically useful reagents. The nucleic acids may also be used in
 CC gene therapy, in chromosome identification, as chromosome markers, or in
 CC generating probes. The PRO polypeptides are useful as molecular markers
 CC for protein electrophoresis, and the isolated nucleic acids may be used
 CC for recombinantly expressing those markers. The PRO polypeptides and
 CC nucleic acids may also be used in tissue typing. Anti-PRO antibodies
 CC are useful in diagnostic assays for PRO, and in affinity purification
 CC of PRO from recombinant cell culture or natural sources. The
 CC present sequence represents a PRO protein.

XX Sequence 382 AA;

Query Match 98.7%; Score 1101; DB 24; Length 382;

Best Local Similarity 96.3%; Pred. No. 9.6e-104;
 Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 ATGRLLS-----QPVCRGGTQPCYKVIVFHTDSRLNFEAEKACRRDGGQLVSI 52
 |||||
 Db 22 ATGRLLSASDLRLRGQGPVCRGGTQPCYKVIVFHTDSRLNFEAEKACRRDGGQLVSI 81
 |||||
 QY 53 ESEDEQKLTEKFTENLLPSDGFWMIGLRREKQSNSTACODLYAMTDGSIOPRNWYVD 112
 |||||
 Db 82 ESEDEQKLTEKFTENLLPSDGFWMIGLRREKQSNSTACODLYAMTDGSIOPRNWYVD 141
 |||||
 QY 113 EPSGSEVCVMYHQPSPAPAGIGPYMFQWDDRCNMKNFICKYSDEKPAVPSREAGE 172
 |||||
 Db 142 EPSGSEVCVMYHQPSPAPAGIGPYMFQWDDRCNMKNFICKYSDEKPAVPSREAGE 201
 |||||
 QY 173 ETELTPVLPEETQBEDAKKTFKESREAAINLAY 206
 |||||
 Db 202 ETELTPVLPEETQBEDAKKTFKESREAAINLAY 235
 |||||

RESULT 14

ABU65578
 ID ABU65578 standard; Protein; 382 AA.

XX AC ABU65578;

XX XX 19-MAY-2003 (first entry)

XX DE Human secreted/transmembrane protein, SEQ ID 20.

XX KW Human; PRO; secreted protein; transmembrane protein;
 KW cytostatic; antiarthritic; osteopathic; adrenal tumour; lung tumour;
 KW colon tumour; breast tumour; prostate tumour; rectal tumour;
 KW cervical tumour; liver tumour; TNF-alpha release; arthritis;
 KW tumour necrosis factor alpha; chondrocyte cell; bone disorder;
 KW cartilage disorder; sports injury.
 XX OS Homo sapiens.

XX PN US2003036156-A1.

XX PD 20-FEB-2003.

XX PF 02-JUL-2002; 2002US-0188767.

XX PR 16-SEP-1998; 98WO-US19330.

PR 07-OCT-1998; 98WO-US21141.

PR 01-DEC-1998; 98WO-US25108.

PR 08-MAR-1999; 99WO-US05028.

PR 14-MAY-1999; 99WO-US10733.

PR 02-JUN-1999; 99WO-US12252.

PR 01-SEP-1999; 99WO-US20111.

PR 15-SEP-1999; 99WO-US21090.

PR 02-DEC-1999; 99WO-US28301.

PR 01-DEC-1999; 99WO-US28551.

PR 30-DEC-1999; 99WO-US31274.

PR 05-JAN-2000; 2000WO-US00219.

PR 18-FEB-2000; 2000WO-US04341.

PR 18-FEB-2000; 2000WO-US04342.

PR 22-FEB-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.

PR 02-MAR-2000; 2000WO-US05841.

PR 15-MAR-2000; 2000WO-US06884.

PR 30-MAR-2000; 2000WO-US08439.

PR 22-MAY-2000; 2000WO-US13705.

PR 17-MAY-2000; 2000WO-US13705.

PR 30-MAY-2000; 2000WO-US14042.

PR 02-JUN-2000; 2000WO-US14941.

PR 26-JUL-2000; 2000WO-US15264.

PR 24-AUG-2000; 2000WO-US20710.

PR 08-NOV-2000; 2000WO-US23328.

PR 01-DEC-2000; 2000WO-US32678.


```
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-098602P.
PR 10-SEP-1998; 98US-099741P.

Query Match 98.7%; Score 1101; DB 24; Length 382;
Best Local Similarity 96.3%; Pred. No. 9.6e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGCTQPCYKVIYFHDTSRLNFEFAKEACRRDGGQLVSI 52
Db |||||
QY 22 ATGRLLSASDLDRGGQVCRGCTQPCYKVIYFHDTSRLNFEFAKEACRRDGGQLVSI 81
Db |||||
QY 53 ESEDEQKLEKFIENLLPSDGFWIGLRREERQSNSTACQDLYAWTDGSIQFRNMYVD 112
Db |||||
QY 82 ESEDEQKLEKFIENLLPSDGFWIGLRREERQSNSTACQDLYAWTDGSIQFRNMYVD 141
Db |||||
QY 113 EPCGSEVCMVHQPSAPAGIGGPFYQWDRCNMKNFICKYSDKPAVPSRAEGE 172
Db |||||
QY 142 EPCGSEVCMVHQPSAPAGIGGPFYQWDRCNMKNFICKYSDKPAVPSRAEGE 201
Db |||||
QY 173 ETELTPVLPEETQEDAKKTFKESRAALNLAY 206
Db |||||
QY 202 ETELTPVLPEETQEDAKKTFKESRAALNLAY 235
Db |||||

RESULT 15
ABU65911
ID ABU65911 standard; Protein; 382 AA.
XX AC ABU65911;
XX DT 20-MAY-2003 (first entry)
XX DE Novel human secreted and transmembrane protein PRO234.
XX KW Human; secreted protein; transmembrane protein; cytostatic;
XX Gene Therapy; TNF-Agonist-Alpha; chondrocyte stimulator; tumour;
XX adrenal tumour; lung tumour; colon tumour; breast tumour;
XX prostate tumour; rectal tumour; cervical tumour; liver tumour.
XX OS Homo sapiens.
XX PN US2003036157-A1.
XX PD 20-FEB-2003.
XX PF 02-JUL-2002; 2002US-0188769.
XX PR 16-SEP-1998; 98WO-US19330.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 08-MAR-1999; 99WO-US05028.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12552.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28551.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00019.
PR 18-FEB-2000; 2000WO-US04341.
PR 19-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.

24-AUG-2000; 2000WO-US23328.
08-NOV-2000; 2000WO-US30952.
01-DEC-2000; 2000WO-US32678.
20-DEC-2000; 2000WO-US34956.
28-FEB-2001; 2001WO-US06520.
01-JUN-2001; 2001WO-US17800.
20-JUN-2001; 2001WO-US19692.
29-JUN-2001; 2001WO-US21066.
09-JUL-2001; 2001WO-US21735.
29-AUG-2001; 2001WO-US27099.
18-SEP-1997; 97US-059263P.
18-SEP-1997; 97US-059266P.
17-OCT-1997; 97US-062250P.
21-OCT-1997; 97US-063486P.
24-OCT-1997; 97US-063120P.
24-OCT-1997; 97US-063121P.
28-OCT-1997; 97US-063540P.
28-OCT-1997; 97US-063541P.
28-OCT-1997; 97US-063544P.
28-OCT-1997; 97US-063564P.
28-OCT-1997; 97US-063734P.
31-OCT-1997; 97US-063870P.
31-OCT-1997; 97US-064103P.
13-NOV-1997; 97US-065311P.
21-NOV-1997; 97US-066120P.
24-NOV-1997; 97US-066466P.
24-NOV-1997; 97US-066772P.
11-DEC-1997; 97US-069335P.
12-DEC-1997; 97US-069425P.
17-DEC-1997; 97US-069870P.
18-DEC-1997; 97US-068017P.
10-MAR-1998; 98US-077450P.
11-MAR-1998; 98US-077632P.
11-MAR-1998; 98US-077649P.
20-MAR-1998; 98US-078886P.
20-MAR-1998; 98US-078939P.
27-MAR-1998; 98US-079664P.
27-MAR-1998; 98US-079786P.
31-MAR-1998; 98US-080107P.
31-MAR-1998; 98US-080194P.
01-APR-1998; 98US-080327P.
01-APR-1998; 98US-080333P.
08-APR-1998; 98US-081049P.
08-APR-1998; 98US-081070P.
09-APR-1998; 98US-081195P.
15-APR-1998; 98US-081838P.
21-APR-1998; 98US-082568P.
21-APR-1998; 98US-082569P.
22-APR-1998; 98US-082704P.
22-APR-1998; 98US-082797P.
28-APR-1998; 98US-083322P.
29-APR-1998; 98US-083495P.
29-APR-1998; 98US-083496P.
29-APR-1998; 98US-083499P.
29-APR-1998; 98US-083559P.
05-MAY-1998; 98US-084366P.
06-MAY-1998; 98US-084414P.
07-MAY-1998; 98US-084639P.
07-MAY-1998; 98US-084640P.
07-MAY-1998; 98US-084643P.
15-MAY-1998; 98US-085579P.
15-MAY-1998; 98US-085580P.
15-MAY-1998; 98US-085582P.
15-MAY-1998; 98US-085700P.
18-MAY-1998; 98US-086023P.
22-MAY-1998; 98US-086392P.
22-MAY-1998; 98US-086486P.
28-MAY-1998; 98US-087098P.
28-MAY-1998; 98US-087208P.
02-JUN-1998; 98US-087609P.
02-JUN-1998; 98US-087759P.
03-JUN-1998; 98US-087827P.
04-JUN-1998; 98US-088025P.
```

PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088036P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090533P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.

PR 01-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.
PR 10-SEP-1998; 98US-099754P.
PR 10-SEP-1998; 98US-099763P.

Query Match 98.7%; Score 1101; DB 24; Length 382;
Best Local Similarity 96.3%; Pred. No. 9.6e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQPCVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLDLRGQPVCRCGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSI 81

Qy 53 ESEDEQKLIKFIENLLPSPDGFWMIGLRRRREEKQSNSTACQDLVYAWTQSGISQPRNWTVD 112
Db 82 ESEDEQKLIKFIENLLPSPDGFWMIGLRRRREEKQSNSTACQDLVYAWTQSGISQPRNWTVD 141

Qy 113 EPSCGSEVCVVMYHQPAPAGIGGYPMPQWNNDRCNMKNFTICKYSDKPAVPSREASGE 172
Db 142 EPSCGSEVCVVMYHQPAPAGIGGYPMPQWNNDRCNMKNFTICKYSDKPAVPSREASGE 201

Qy 173 ETELTTPVLPEETOEDAKKTFKESREAAALNLAY 206
Db 202 ETELTTPVLPEETOEDAKKTFKESREAAALNLAY 235

Search completed: December 22, 2003, 16:10:07
Job time : 30.1819 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:05:39 ; Search time 43.5759 Seconds
(without alignments)
2214.797 Million cell updates/sec

Title: US-09-887-855-2

Perfect score: 2000

Sequence: 1 MRPGTALQAVLLAVLLVGLR.....PDQMGRSKESGWENEIYGY 374

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phage:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_rvirus:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|---------------------|
| 1 | 2000 | 100.0 | 374 | 4 Q96NF3 | Q96nf3 homo sapien |
| 2 | 1997 | 99.9 | 374 | 4 Q8TAY8 | Q8tay8 homo sapien |
| 3 | 1996 | 99.8 | 374 | 4 Q96NC5 | Q96nc5 homo sapien |
| 4 | 1697.5 | 84.9 | 374 | 11 Q92209 | Q92209 cricetus |
| 5 | 862.5 | 43.1 | 211 | 11 Q8C351 | Q8c351 mus musculus |
| 6 | 652 | 32.6 | 292 | 11 Q8BVU2 | Q8bv2 mus musculus |
| 7 | 646 | 32.3 | 246 | 11 Q8BM17 | Q8bmi7 mus musculus |
| 8 | 185 | 9.2 | 1290 | 13 Q9W6E1 | Q9w6e1 gallus gall |
| 9 | 178.5 | 8.9 | 1456 | 11 Q61830 | Q61830 mus musculus |
| 10 | 177.5 | 8.9 | 1348 | 5 Q25199 | Q25199 hydra atten |
| 11 | 176.5 | 8.8 | 1479 | 4 Q9Y3P9 | Q9y3p9 homo sapien |
| 12 | 176.5 | 8.8 | 1479 | 4 Q9UBG0 | Q9ubg0 homo sapien |
| 13 | 170.5 | 8.5 | 1479 | 11 Q64449 | Q64449 mus musculus |
| 14 | 169 | 8.5 | 217 | 11 Q8C4F8 | Q8c4f8 mus musculus |
| 15 | 165.5 | 8.3 | 742 | 11 Q8K4Q8 | Q8k4q8 mus musculus |
| 16 | 165.5 | 8.3 | 742 | 11 Q8C979 | Q8c979 mus musculus |

| | | | | | |
|----|-------|-----|------|-----------|---------------------|
| 17 | 161.5 | 8.1 | 134 | 5 Q9XYX3 | Q9xyx3 hydra magni |
| 18 | 159.5 | 8.0 | 742 | 11 Q8VIF6 | Q8vif6 mus musculus |
| 19 | 159 | 8.0 | 1031 | 5 Q8WSX2 | Q8wsx2 dugesia tig |
| 20 | 158.5 | 7.9 | 719 | 6 Q8Z623 | Q8z623 bos taurus |
| 21 | 158 | 7.9 | 142 | 11 Q8CJ86 | Q8cj86 mus musculus |
| 22 | 158 | 7.9 | 142 | 11 Q8BHK7 | Q8bhk7 mus musculus |
| 23 | 158 | 7.9 | 295 | 11 Q91ZM4 | Q91zm4 mus musculus |
| 24 | 158 | 7.9 | 311 | 11 Q9D8V4 | Q9d8v4 mus musculus |
| 25 | 158 | 7.9 | 325 | 11 Q91ZX0 | Q91zx0 mus musculus |
| 26 | 157.5 | 7.9 | 358 | 6 Q8HY04 | Q8hy04 papio hamad |
| 27 | 157.5 | 7.9 | 381 | 6 Q8SQB2 | Q8sqb2 macaca mula |
| 28 | 157.5 | 7.9 | 652 | 4 Q8IXK1 | Q8ixk1 homo sapien |
| 29 | 157 | 7.8 | 158 | 13 Q90WI7 | Q90wi7 bungarus fa |
| 30 | 156 | 7.8 | 162 | 5 Q25459 | Q25459 megabalanus |
| 31 | 156 | 7.8 | 323 | 11 Q8CJ91 | Q8cj91 mus musculus |
| 32 | 156 | 7.8 | 339 | 6 Q95244 | Q95244 sus scrofa |
| 33 | 155 | 7.8 | 293 | 11 Q8BGZ0 | Q8bgz0 mus musculus |
| 34 | 155 | 7.8 | 323 | 11 Q8CJ94 | Q8cj94 mus musculus |
| 35 | 155 | 7.8 | 323 | 11 Q8CJ93 | Q8cj93 mus musculus |
| 36 | 155 | 7.8 | 323 | 11 Q8CJ88 | Q8cj88 mus musculus |
| 37 | 155 | 7.8 | 381 | 6 Q95LA8 | Q95la8 macaca mula |
| 38 | 155 | 7.8 | 1152 | 13 Q90WM2 | Q90wm2 xenopus lae |
| 39 | 154 | 7.7 | 312 | 6 Q8HXL6 | Q8hxl6 macaca mula |
| 40 | 154 | 7.7 | 322 | 11 Q8CJ89 | Q8cj89 mus musculus |
| 41 | 154 | 7.7 | 323 | 11 Q8CJ92 | Q8cj92 mus musculus |
| 42 | 154 | 7.7 | 381 | 6 Q95LC6 | Q95lc6 macaca neme |
| 43 | 154 | 7.7 | 381 | 6 Q8HXL7 | Q8hxl7 macaca mula |
| 44 | 154 | 7.7 | 404 | 6 Q95J96 | Q95j96 macaca mula |
| 45 | 152.5 | 7.6 | 427 | 6 Q8HYB9 | Q8hyb9 pan troglod |

ALIGNMENTS

RESULT 1

Q96NF3 PRELIMINARY; PRT; 374 AA.

AC Q96NF3; 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ30977.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie K., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Negai K., Isogai T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK055539; BAB70946.1; .
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT_1.
DR PROSITE; PS0041; C_TYPE_LLECTIN_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 374 AA; 42280 MW; 8AB64E6BC9E56DCD CRC64;

Query Match 100.0%; Score 2000; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.5e-180;
Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLSGQPCVCRGCTGTPCYKVIYFHDTSRLNFEAK 60
|||||
DB 1 MRPGTALQAVLLAVLLVGLRAATGRLLSGQPCVCRGCTGTPCYKVIYFHDTSRLNFEAK 60
|||||

QY 61 EACRRDGGQLVSIESTEDEOKLIEKFIENLLPSDGFWMGLRRREEKQSNSTACQDLYAWT 120
DB 61 EACRRDGGQLVSIESTEDEOKLIEKFIENLLPSDGFWMGLRRREEKQSNSTACQDLYAWT 120
QY 121 DGSISQFRNYYVDEPSCGSEVVMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSD 180
DB 121 DGSISQFRNYYVDEPSCGSEVVMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSD 180
QY 181 EKPAVPSREAEGETELTTPVLPETQEDAKKTFKESREAAALNAYILIPSIPLLLLV 240
DB 181 EKPAVPSREAEGETELTTPVLPETQEDAKKTFKESREAAALNAYILIPSIPLLLLV 240
QY 241 VTTVVCWVWICRKRREQDPSTKKQHTIWPSPHQNSPDLEVYVIRKQSEADLAETRP 300
DB 241 VTTVVCWVWICRKRREQDPSTKKQHTIWPSPHQNSPDLEVYVIRKQSEADLAETRP 300
QY 301 DLKNIISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYEFSPQOMGR 360
DB 301 DLKNIISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYEFSPQOMGR 360
QY 361 SKESGWVENEIYGY 374
DB 361 SKESGWVENEIYGY 374
RESULT 2
Q8TAY8
ID Q8TAY8 PRELIMINARY; PRT; 374 AA.
AC Q8TAY8
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Similar to unnamed protein product.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025407; AAH25407.1; -.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PSS0041; C TYPE LECTIN 2; 1.
SQ SEQUENCE 374 AA; 42312 MW; FC214E6BC9E578D9 CRC64;
Query Match 99.9%; Score 1997; DB 4; Length 374;
Best Local Similarity 99.7%; Pred. No. 2.9e-180;
Matches 373; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLSQPVCRGQTPCYKVIYFHDTSRLNFEAK 60
DB 1 MRPGTALQAVLLAVLLVGLRAATGRLLSQPVCRGQTPCYKVIYFHDTSRLNFEAK 60
QY 61 EACRRDGGQLVSIESTEDEOKLIEKFIENLLPSDGFWMGLRRREEKQSNSTACQDLYAWT 120
DB 61 EACRRDGGQLVSIESTEDEOKLIEKFIENLLPSDGFWMGLRRREEKQSNSTACQDLYAWT 120
QY 121 DGSISQFRNYYVDEPSCGSEVVMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSD 180
DB 121 DGSISQFRNYYVDEPSCGSEVVMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSD 180
QY 181 EKPAVPSREAEGETELTTPVLPETQEDAKKTFKESREAAALNAYILIPSIPLLLLV 240
DB 181 EKPAVPSREAEGETELTTPVLPETQEDAKKTFKESREAAALNAYILIPSIPLLLLV 240
QY 241 VTTVVCWVWICRKRREQDPSTKKQHTIWPSPHQNSPDLEVYVIRKQSEADLAETRP 300
DB 241 VTTVVCWVWICRKRREQDPSTKKQHTIWPSPHQNSPDLEVYVIRKQSEADLAETRP 300

QY 301 DLKNIISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYEFSPQOMGR 360
DB 301 DLKNIISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYEFSPQOMGR 360
QY 361 SKESGWVENEIYGY 374
DB 361 SKESGWVENEIYGY 374
RESULT 3
Q96NC5
ID Q96NC5 PRELIMINARY; PRT; 374 AA.
AC Q96NC5
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ31092.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K.,
RA Isogai T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK055654; BAB70978.1; -.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PSS0041; C TYPE LECTIN 2; 1.
KW Hypothetical protein.
SQ SEQUENCE 374 AA; 42310 MW; CBF74E676E23BA16 CRC64;
Query Match 99.8%; Score 1996; DB 4; Length 374;
Best Local Similarity 99.7%; Pred. No. 3.7e-180;
Matches 373; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLSQPVCRGQTPCYKVIYFHDTSRLNFEAK 60
DB 1 MRPGTALQAVLLAVLLVGLRAATGRLLSQPVCRGQTPCYKVIYFHDTSRLNFEAK 60
QY 61 EACRRDGGQLVSIESTEDEOKLIEKFIENLLPSDGFWMGLRRREEKQSNSTACQDLYAWT 120
DB 61 EACRRDGGQLVSIESTEDEOKLIEKFIENLLPSDGFWMGLRRREEKQSNSTACQDLYAWT 120
QY 121 DGSISQFRNYYVDEPSCGSEVVMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSD 180
DB 121 DGSISQFRNYYVDEPSCGSEVVMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSD 180
QY 181 EKPAVPSREAEGETELTTPVLPETQEDAKKTFKESREAAALNAYILIPSIPLLLLV 240
DB 181 EKPAVPSREAEGETELTTPVLPETQEDAKKTFKESREAAALNAYILIPSIPLLLLV 240
QY 241 VTTVVCWVWICRKRREQDPSTKKQHTIWPSPHQNSPDLEVYVIRKQSEADLAETRP 300
DB 241 VTTVVCWVWICRKRREQDPSTKKQHTIWPSPHQNSPDLEVYVIRKQSEADLAETRP 300
QY 301 DLKNIISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYEFSPQOMGR 360
DB 301 DLKNIISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYEFSPQOMGR 360
QY 361 SKESGWVENEIYGY 374
DB 361 SKESGWVENEIYGY 374

```

RESULT 4
Q92209 PRELIMINARY; PRT; 374 AA.
AC Q92209;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Laylin.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RN SEQUENCE FROM N.A.
RA Borowsky M.L., Hynes R.O.;
RT "Laylin, a novel talin-binding transmembrane protein homologous with
RT C-type lectins, is localized in membrane ruffles.";
RL J. Cell Biol. 143:0-0(1998)
DR EMBL; AF093673; AAC68695.1; -- 132
DR HSP; P22897; 1EGG.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS0041; C TYPE LECTIN 2; 1.
SQ SEQUENCE 374 AA; 42435 MW; 298A8BA24FB04E1C CRC64;

Query Match 84.9%; Score 1697.5; DB 11; Length 374;
Best Local Similarity 84.7%; Pred. No. 6.2e-152;
Matches 316; Conservative 23; Mismatches 33; Indels 1; Gaps 1;

Qy 1 MRPGTALQAVLLAVLLVGLRAATGRLLSQQVCRGGTORPCYKVIYFHDTSRRLNFEAK 60
Db 1 MQPGPALQAVLLAVLLVGLSEPRSSKGRLLSGQLVCRGGTRPCYKVIYFHDFAQRNFEEAK 60

Qy 61 EACRRDGGQLVSEDEQKLIETKFIENLLPSDGFWMGLRRRREKQSNSTACODLYAWT 120
Db 61 EACRRDGGQLVSEDEQKLIETKFIENLLASDGFWMGLRRLEVKQVNTACQDLYAWT 120

Qy 121 DGSISQFRNMYVDEPSCGSEVCVMYHQPAPAGIGPYFQWDDRCNKNKFNICKYSD 180
Db 121 DGSISQFRNMYVDEPSCGSEVCVMYHQPAPAGIGGYSYFQWDDRCNKNKFNICKYAD 180

Qy 181 EKPA-VPSREAEGETELTTPVLPETQEDAKTKFESREAAINLAYILIPILLLL 239
Db 181 EKPSITPSIRPGGEATEPTPLPETQKEDTKETKESREAAINLAYILIPILPLL 240

Qy 240 VTTTVCWVWICRRKREQDPTTKQHTIWPSPHQNSPDLEVYVIRKQSEADLAETR 299
Db 241 VVTSAAACWVICRRRKQEQDPTTKQHTIWPSPHQNSPNLDVYVIRKQSEADLTEPR 300

Qy 300 PDLKNIISFRVCSGEATPDDMSCDYDNMNVNPSSEGVTLVSVSGFVTNDIYFSPDMG 359
Db 301 PDLKNIISFRVCSGEATPDDMSCDYDNMNVNPSSEGVTLVSVSGFVTNDIYFSPDMG 360

Qy 360 RSKESGVNEIY 372
Db 361 RSKESGVNEIY 373

RESULT 5
Q8C351 PRELIMINARY; PRT; 211 AA.
AC Q8C351;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE LAYLIN homolog (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK086930; BAC39765.1; --
FT NON TER 211
SQ SEQUENCE 211 AA; 23697 MW; AD9870B5957DD5AE CRC64;

Query Match 43.1%; Score 862.5; DB 11; Length 211;
Best Local Similarity 77.7%; Pred. No. 2.6e-73;
Matches 164; Conservative 12; Mismatches 26; Indels 9; Gaps 2;

Qy 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGTORPCYKVIYFHDTSR 52
Db 1 MQPGPALQAVLLAVLLVGLAKPRDSKGRLLSASDLDPGGQLVCRGGTRPCYKVIYFHDFAQ 60

Qy 53 RLNFEEAKEACRRDGGQLVSEDEQKLIETKFIENLLPSDGFWMGLRRRREKQSNSTA 112
Db 61 RLNFEEAKEACRRDGGQLVSEDEQKLIETKFIENLLASDGFWMGLRRRREKQSNNTA 120

Qy 113 CODLYAWTDGSIQFRNMYVDEPSCGSEVCVMYHQPAPAGIGPYFQWDDRCNKNK 172
Db 121 CODLYAWTDGSIQFRNMYVDEPSCGSEVCVMYHQPAPAGIGGYSYFQWDDRCNKNK 180

Qy 173 NFICKYSDEKPA-VPSREAEGETELTTPVL 202
Db 181 NFICKYHDDKPSITPTSPWPGGEATEPATPLL 211

RESULT 6
Q8BVU2 PRELIMINARY; PRT; 292 AA.
AC Q8BVU2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE C-type lectin protein MT75 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK076523; BAC36378.1; --
SQ SEQUENCE 292 AA; 32502 MW; 73B631C0714D54E2 CRC64;

Query Match 32.6%; Score 652; DB 11; Length 292;
Best Local Similarity 47.0%; Pred. No. 3.2e-53;
Matches 140; Conservative 39; Mismatches 79; Indels 40; Gaps 9;

Qy 10 VLLAVLLVGLRAATGRLLSGQVCRGGTORPCYKVIYFHDTSRRLNFEAKEACRRDGGQ 69
Db 8 LLGAALLCAGAFAREWSGQVCFADVKHPCYKVIYFHDTSRRLNFEAKEACRRDGGQ 67

Qy 70 LVSISEDEQKLIETKFIENLLP-----SDGDFWMGLRRRREKQSNSTACODLYAWTDGSI 124
Db 68 LLSLENEAEQKLIETKFIENLLP-----SDGDFWMGLRRRREKQSNSTACODLYAWTDGSI 126

Qy 125 SOFRNMYVDEPSCGSEVCVMYHQPAPAGIGPYFQWDDRCNKNKFNICKYSD-KP 183

```

```
Db 127 SQFRNWTDBPCSGSEKCVVMYHQTANPGLGPLYQWDDRCNMKNHNYCKYEPIHP 186
Qy 184 AVPSREAEGETELTPVLPPEETOEDAKTKFKESREAA--NLAYILIPSTPLLLLVV 241
Db 187 TEPA-----EKPLYTNQ--PEETHENV-----VTEAGIIPNLIVYIIPITPLLLLV 233
Qy 242 TTVVCVWVICRKR-----EQDPSTKK-----OHTWPSPHQNSPDL 281
Db 234 ALGTCCFQMLHKKARRHFHKDSTPLSSSECLAESNLVHMAGSLIPYHFQNNSPSL 291

RESULT 7
Q8BM17 PRELIMINARY; PRT; 246 AA.
AC Q8BM17;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE C-type lectin protein MT75 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Forelimb;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RL EMBL; AK031063; BAC27234.1; -.
SQ SEQUENCE 246 AA; 27394 MW; E872660C58267752 CRC64;

Query Match 32.3%; Score 646; DB 11; Length 246;
Best Local Similarity 52.2%; Pred. No. 9.5e-53;
Matches 132; Conservative 34; Mismatches 65; Indels 22; Gaps 7;

Qy 10 VLLAVLLVGLRAATGRLLSQPVCRGQTQPCVKYVYEHDTSRLLNFEAKEACRDGQ 69
Db 8 LIGAAULLCAGQAFARVSVGQKVCFADVRHPCYKMAFYFHELSRVSFQEARLACESEGV 67
Qy 70 LVSISEDEOKLIEKFIENLLP-----SDGDFWIGLRREEKQNSTACQDLYAWTDGSI 124
Db 68 LLSLENAEQKLIESMLQNLTKPTGTISDGFWIGLRSDGQT-SGAPDLYQWSDGSS 126
Qy 125 SQFRNWTDBPCSGSEKCVVMYHQTANPGLGPLYQWDDRCNMKNHNYCKYEPIHP 183
Db 127 SQFRNWTDBPCSGSEKCVVMYHQTANPGLGPLYQWDDRCNMKNHNYCKYEPIHP 186
Qy 184 AVPSREAEGETELTPVLPPEETOEDAKTKFKESREAA--NLAYILIPSTPLLLLVV 241
Db 187 TEPA-----EKPLYTNQ--PEETHENV-----VTEAGIIPNLIVYIIPITPLLLLV 233
Qy 242 TTVVCVWVICRKR 254
Db 234 ALGTCCFQMLHKK 246

RESULT 8
Q9W6E1 PRELIMINARY; PRT; 1290 AA.
AC Q9W6E1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Neurocan core protein.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=20309833; PubMed=10951024;
Li H., Leung T.C., Hoffman S., Balsamo J., Lilien J.;
RT "Coordinate Regulation of Cadherin and Integrin Function by the
RT Chondroitin Sulfate Proteoglycan Neurocan."
RL J. Cell Biol. 149:1275-1288(2000).
DR EMBL; AF116856; AAD24546.2; -.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca..
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR003599; Ig_.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR0010; EGPBLOOD.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS01241; Link; 2.
DR EGF-like domain.
SQ SEQUENCE 1290 AA; 138877 MW; 182BD86D0E40BE78 CRC64;

Query Match 9.2%; Score 185; DB 13; Length 1290;
Best Local Similarity 34.7%; Pred. No. 3.3e-08;
Matches 50; Conservative 18; Mismatches 42; Indels 34; Gaps 8;

Qy 38 QRPCYKVIYFHDTSRLNFEAKEACRDGQGLVSISEDEOKLIEKFIENLLPSDGF 97
Db 1064 QGHCYR--YF---SRRSWEDAERDCRRRAGHLTSHSQEHCPINSF-----GHENTW 1112
Qy 98 IGLRREEKQNSTACQDLYAWTDGSIQFRNWTDEPS---CGSEVCVVMY-HQPSAPA 153
Db 1113 IGLNDRIVQD-----FQWTDNTGLQYENWRENQPNPFAGGDCVVLVSHE----- 1159
Qy 154 GIGGPFYMFQWDDRCNMKNFICK 177
Db 1160 -IG-----KWNQVPCNYPICK 1177

RESULT 9
Q61830 PRELIMINARY; PRT; 1456 AA.
ID Q61830;
AC Q61830;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Macrophage mannose receptor precursor.
GN MRC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=93043353; PubMed=1421407;
RA Harris N., Rice M., Chang G., Ezekowitz R.B.;
RT "Characterization of the murine macrophage mannose receptor.";
RL Blood 80:2363-2373(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Super M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z11974; CAA78028.1; -.
DR HSSP; P22897; 1EGG.
DR MGD; MGI:971142; Mrc1.
DR InterPro; IPR002353; AntifreezeII.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin_c; 8.
DR Pfam; PF00652; Ricin_B_lectin; 2.
DR PRINTS; PR00356; ANTIFREEZEII.
DR PRINTS; PR00013; FNTYPEII.
DR ProDom; PD000995; FN_Type_II; 1.
DR SMART; SM00034; CLECT; 8.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 6.
DR PROSITE; PS00441; C_TYPE_LECTIN_2; 8.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS02331; RICIN_B_LECTIN; 1.
KW Receptor; Signal.
FT SIGNAL
FT CHAIN
FT CHAIN
SQ SEQUENCE 1456 AA; 165065 MW; 4EBD3F1B8619A594 CRC64;

Query Match 8.9%; Score 178.5; DB 11; Length 1456;
Best Local Similarity 25.9%; Pred. No. 1.6e-07;
Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;

Qy 42 YKVIYFHDTSRLNFEAKEACRRDGGQGVLSIESDEOKLIEKFENLPSDGDFTWGLR 101
Db 807 YKDYQYFFSKEKETMDNARFCKNFGDLATIKSEBKFLWKYI-NKNGQSPYFGLM 865
Qy 102 RREKQSNSTACQDLYAWTQGISQERNWYVDEPSCGS--EVCVVMYHQPSAPAGIGPY 159
Db 866 ISMDKK-----FIWMDGSKVDVFAWATGEFANFDDNCVTMY-----TNSGF 908
Qy 160 MFQWDDRCNMKNFNICK---YSDEKPAVPSREAEGETELTTPVLPEETQE----- 208
Db 909 ---WINDINGVYPNNFICORHNSINATAMP-----TTPTPGCKEGHLYKNK 954
Qy 209 -----EAKTKFESREALNL 225
Db 955 CFKIFGPANBEKSWQDARQCKGL 979

RESULT 10
ID Q25199 PRELIMINARY; PRT; 1348 AA.
AC Q25199;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Tyrosine kinase receptor.
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Anthomedusae;
OC Hydrozoa; Hydra.
OX NCBI_TaxID=6087;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=Irvine;
RX MEDLINE=20209407; PubMed=10744720;
RA Reidling J.C., Miller M.A., Steele R.E.;
RT "Sweet tooth, a Novel Receptor Protein-Tyrosine Kinase with C-type Lectin-like Extracellular Domains.";
RL J. Biol. Chem. 275:10323-10330(2000).
DR EMBL; L22612; AAA29218.2; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR003990; Pancreatias_ac.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00059; lectin_c; 4.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR01504; PNCREATITSAP.
DR ProDom; PD000001; Prot_kinase; 2.
DR SMART; SM00034; CLECT; 4.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 2.
DR PROSITE; PS00441; C_TYPE_LECTIN_2; 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1348 AA; 156916 MW; 77D2122093227FEF CRC64;

Query Match 8.9%; Score 177.5; DB 5; Length 1348;
Best Local Similarity 28.0%; Pred. No. 1.8e-07;
Matches 47; Conservative 32; Mismatches 46; Indels 43; Gaps 8;

Qy 27 LSGQPVG--RGGTQRPC-----YKVIYFHDTSRLNFEAKEACRRDGGQGVLSIESDE 78
Db 412 LSHRFICKVKRATNEYCAEGWTSYRYCYFYISYEFDFWFKSFSSCONIGNLLSTENDEE 471
Qy 79 OKLIEKFENLPSDGD-FWIGLRR-----RREKQSNSTACQDLYAWTQGISQERNWY 131
Db 472 ---NREIENDLIKNDKYNIGLNKWNLYKNKR-----FEWSDNTYTFQFNWI 518
Qy 132 VDEP--SCGSEVCVVMYHQPSAPAGIGPYFQWDDRCNMKNFNICK 177
Db 519 TNQPDNNNGIESCVEMNYN-----GMSDECKEYLVNGFICK 553

RESULT 11
ID Q9Y5P9 PRELIMINARY; PRT; 1479 AA.
AC Q9Y5P9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Endocytic receptor Endo180.
GN ENDO180.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20148849; PubMed=10683150;
RA Sheikh H., Yarwood H., Ashworth A., Isaack C.;
RT "Endo180, an endocytic recycling glycoprotein related to the macrophage mannose receptor is expressed on fibroblasts, endothelial cells and macrophages and functions as a lectin receptor.";
RL J. Cell Sci. 113:1021-1032(2000).
DR EMBL; AF134838; AAD30280.1; -.
DR HSSP; P02751; 2FN2.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000566; Lipocln_cytFASP.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin_c; 8.
DR PRINTS; PR00013; FNTYPEII.

```


DR ProDom; PD000995; FN Type II; 1.
 DR SMART; SM00034; CLECT; 8.
 DR SMART; SM00059; FN2; 1.
 DR SMART; SM00458; RICIN; 1.
 DR PROSITE; PS00615; C_TYPE_LECTIN_1; 3.
 DR PROSITE; PS00041; C_TYPE_LECTIN_2; 8.
 DR PROSITE; PS00023; FIBONECTIN_2; 1.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 DR PROSITE; PS00231; RICIN_B_LECTIN; 1.
 KW Receptor.
 SQ SEQUENCE 1479 AA; 166669 MW; 9F4BAF355F036FCE CRC64;
 Query Match 8.8%; Score 176.5; DB 4; Length 1479;
 Best Local Similarity 30.4%; Pred. No. 2.6e-07;
 Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;
 QY 28 SGQVCRGGTQRCYKVIYFHDTSRRINLFEAEACRRDGGQLVSISEDEQKLEKFI 87
 DB 385 SWQPP-----QGHCYRL-----QAEKRSWQESKKACLRGGDLVSIHMAELEFITKQIK 434
 QY 88 NLLPSDGDWFGLRRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCVV 144
 DB 435 QEVE---ELWIGL-----NDLKQMFENSDGSLVSTHWHPPFPNNFRSLDCV 483
 QY 145 MYHOPSAPAGIGGYPYFQWDDRCNNKNNFKICKYDEKPAVPSREAE 192
 DB 484 IW-----GPEG-----RWNDSPCNQSLPSICKKAGQLSQGAAEDHG 520

RESULT 12

Q9UBG0
 ID Q9UBG0 PRELIMINARY; PRT; 1479 AA.
 AC Q9UBG0;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Urokinase receptor-associated protein UPARAP.
 GN KIAA0709.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Behrendt N., Jensen O.N., Engelholm L.H., Mortz E., Mann M., Dano K.,
 RT "A urokinase receptor-associated protein with specific collagen-
 RT binding properties".
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98403880; PubMed=9734811;
 RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro".
 RL DNA Res. 5:169-176(1998).
 DR EMBL; AF107292; AAF14192.1; -.
 DR EMBL; AB014609; BAA31684.1; -.
 DR HSSP; P02751; 2FN2.
 DR InterPro; IPR000562; FN_Type_II.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000566; Lipocln_cytFABP.
 DR Pfam; PF00040; fn2; 1.
 DR Pfam; PF00059; lectin_c; 8.
 DR PRINTS; PR00013; FNTYPEII.
 DR ProDom; PD000995; FN_Type_II; 1.
 DR SMART; SM00034; CLECT; 8.
 DR SMART; SM00059; FN2; 1.
 DR SMART; SM00458; RICIN; 1.
 DR PROSITE; PS00615; C_TYPE_LECTIN_1; 3.
 DR PROSITE; PS00023; FIBONECTIN_2; 1.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 KW Receptor.
 SQ SEQUENCE 1479 AA; 166669 MW; 9F4BAF355F036FCE CRC64;
 Query Match 8.8%; Score 176.5; DB 4; Length 1479;
 Best Local Similarity 30.4%; Pred. No. 2.6e-07;
 Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;
 QY 28 SGQVCRGGTQRCYKVIYFHDTSRRINLFEAEACRRDGGQLVSISEDEQKLEKFI 87
 DB 385 SWQPP-----QGHCYRL-----QAEKRSWQESKKACLRGGDLVSIHMAELEFITKQIK 434
 QY 88 NLLPSDGDWFGLRRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCVV 144
 DB 435 QEVE---ELWIGL-----NDLKQMFENSDGSLVSTHWHPPFPNNFRSLDCV 483
 QY 145 MYHOPSAPAGIGGYPYFQWDDRCNNKNNFKICKYDEKPAVPSREAE 192
 DB 484 IW-----GPEG-----RWNDSPCNQSLPSICKKAGQLSQGAAEDHG 520

DR PROSITE; PS00041; C_TYPE_LECTIN_2; 8.
 DR PROSITE; PS00023; FIBONECTIN_2; 1.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 DR PROSITE; PS00231; RICIN_B_LECTIN; 1.
 KW Kinase; Receptor.
 SQ SEQUENCE 1479 AA; 166654 MW; C7583EA78E2792D1 CRC64;
 Query Match 8.8%; Score 176.5; DB 4; Length 1479;
 Best Local Similarity 30.4%; Pred. No. 2.6e-07;
 Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;
 QY 28 SGQVCRGGTQRCYKVIYFHDTSRRINLFEAEACRRDGGQLVSISEDEQKLEKFI 87
 DB 385 SWQPP-----QGHCYRL-----QAEKRSWQESKKACLRGGDLVSIHMAELEFITKQIK 434
 QY 88 NLLPSDGDWFGLRRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCVV 144
 DB 435 QEVE---ELWIGL-----NDLKQMFENSDGSLVSTHWHPPFPNNFRSLDCV 483
 QY 145 MYHOPSAPAGIGGYPYFQWDDRCNNKNNFKICKYDEKPAVPSREAE 192
 DB 484 IW-----GPEG-----RWNDSPCNQSLPSICKKAGQLSQGAAEDHG 520

RESULT 13

Q64449
 ID Q64449 PRELIMINARY; PRT; 1479 AA.
 AC Q64449;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Lectin lambda.
 GN MRC2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96355501; PubMed=8702911;
 RA Wu K., Yuan J., Lasky L.A.;
 RT "Characterization of a novel member of the macrophage mannose receptor
 RT type C lectin family".
 RL J. Biol. Chem. 271:21323-21330(1996).
 DR EMBL; U56734; AAC52729.1; -.
 DR HSSP; P02751; 2FN2.
 DR MGD; MGI:107818; Mrc2.
 DR InterPro; IPR002353; AntifreezeII.
 DR InterPro; IPR001128; Cytochrome P450.
 DR InterPro; IPR000562; FN_Type_II.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000566; Lipocln_cytFABP.
 DR InterPro; IPR000772; Ricin_B_lectin.
 DR Pfam; PF00040; fn2; 1.
 DR Pfam; PF00059; lectin_c; 8.
 DR PRINTS; PR00356; ANTIFREEZEII.
 DR PRINTS; PR00013; FNTYPEII.
 DR ProDom; PD000995; FN_Type_II; 1.
 DR SMART; SM00034; CLECT; 8.
 DR SMART; SM00059; FN2; 1.
 DR SMART; SM00458; RICIN; 1.
 DR PROSITE; PS00086; CYTOCHROME P450; 1.
 DR PROSITE; PS00615; C_TYPE_LECTIN_1; 3.
 DR PROSITE; PS00041; C_TYPE_LECTIN_2; 8.
 DR PROSITE; PS00023; FIBONECTIN_2; 1.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 DR PROSITE; PS00231; RICIN_B_LECTIN; 1.
 SQ SEQUENCE 1479 AA; 167112 MW; 62D456E1089B48C1 CRC64;
 Query Match 8.5%; Score 170.5; DB 11; Length 1479;
 Best Local Similarity 31.4%; Pred. No. 9.4e-07;
 Matches 48; Conservative 22; Mismatches 48; Indels 35; Gaps 7;

```
Qy 28 SQPVCRRGRTQPCYKVIYFHTDTSRLNFEAKACRRDGGQLVSIIEDEQKLIKIEFIE 87
Db 384 SWQPF-----QCHCYRL-----QAERKSWQSKRACLGGGDLLSIHSMAELEFITKQIK 433
Qy 88 NLLPSDGFDTGLRRREKQSNSTACQDLYAWTDSISQFRNMYVDEPS-----CGSEVCVV 144
Db 434 QEVE---ELWIGL-----NDLKQWNPFWSDGSLVSTFTHMHPFPFNFRDLSLDCVT 482
Qy 145 MYHQSAPAGIGGPFQWMDRCNMKNFKICK 177
Db 483 IW----GPEG-----RWNDSPCNQSLPSICK 504

RESULT 14
Q8C4F8 PRELIMINARY; PRT; 217 AA.
AC Q8C4F8:
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Chondroitin sulfate proteoglycan 3 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).
RL EMBL: AK082298; BAC38458.1; -.
FT NON TER 1
SQ SEQUENCE 217 AA; 25949 MW; 5F1A6A57B05B76D6 CRC64;

Query Match 8.5%; Score 169; DB 11; Length 217;
Best Local Similarity 31.1%; Pred. No. 1e-07;
Matches 42; Conservative 16; Mismatches 47; Indels 30; Gaps 5;

Qy 46 YFHTDTSRLNFEAKACRRDGGQLVSIIEDEQKLIKIEFIENLLPSDGFWGLRRREE 105
Db 3 YF---AHRRAWEDAERDCRRAGHLTSVSPSEHKFINF-----GHENSWIGLNDRTV 53
Qy 106 QKSNSTACQDLYAWTDSISQFRNMYVDEPS---CGSEVCVMYHQSAPAGIGGPFQ 162
Db 54 ERD-----FQWTDNTGLQYENWREKQPDNFFAGDCVVMVHESG-----R 95
Qy 163 WNDRCNMKNFKICK 177
Db 96 WNDVPCNYNLPYVCK 110

RESULT 15
Q8K4Q8 PRELIMINARY; PRT; 742 AA.
AC Q8K4Q8:
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Collectin placenta 1.
GN COLEC12 OR CL-P1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtani K., Suzuki Y., Eda S., Kawai T., Kase T., Keshi H., Sakai Y.,
RA Fukuh A., Sakamoto T., Itabe H., Suzutani T., Ogasawara M.,
RA Yoshida I., Wakamiya N.;
```

```
*cDNA cloning of mouse CL-P1 gene.*;
Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
EMBL: AB078434; BAC05523.1; -.
MGD: MGI.2152907; Colecl2.
DR InterPro; IPR002353; AntifreezeII.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; Lectin C.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF00059; lectin C; 1.
DR PRINTS; PR00356; ANTI-FREEZEII.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
DR PROSITE; PS0041; C-TYPE LECTIN 2; 1.
SQ SEQUENCE 742 AA; 81304 MW; 1537C490E5911C45 CRC64;
```

```
Query Match 8.3%; Score 165.5; DB 11; Length 742;
Best Local Similarity 33.8%; Pred. No. 1.1e-06;
Matches 50; Conservative 17; Mismatches 54; Indels 27; Gaps 8;

Qy 41 CYKVIYFHTDTSRLNFEAKACRRDGGQLVSIIEDEQKLIKIEFIENLLPSDGFWGL 100
Db 618 CY---YF---SLEKEIFEDAKLFCEDKSSHLVFINREEQWIKKH-----TVGRSHWIGL 668
Qy 101 RRREEKQSNSTACQDLYAWTDSISQFRNMYVDEP-SCGSEVCVMYHQSAPAGIGG 159
Db 669 TDSEQSE-----WKWLDGSPVDYKNWKAGQPDNWGSG-----HGPGEDCA-GLIY 713
Qy 160 MFQWDDRCNMKNFKICKYDEKPAVPS 187
Db 714 AGQWDFQCDSEINNFI--KEREAVPS 739
```

Search completed: December 22, 2003, 16:13:29
Job time : 45.7759 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:05:39 ; Search time 24.0017 Seconds
(without alignments)
2214.797 Million cell updates/sec

Title: US-09-887-855-5

Perfect score: 1115

Sequence: 1 ATGRLSGQPVCRGGTQRPC.....EEDAKKTFKREDAALNLAY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:*

1: sp_archea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description |
|------------|-------|---------------|--------|--------|---------------------|
| 1 | 1115 | 100.0 | 374 | Q8TAY8 | Q8TAY8 homo sapien |
| 2 | 1115 | 100.0 | 374 | Q96NF3 | Q96NF3 homo sapien |
| 3 | 1111 | 99.6 | 374 | Q96NC5 | Q96NC5 homo sapien |
| 4 | 945.5 | 84.8 | 374 | Q92209 | Q92209 cricetus |
| 5 | 803.5 | 72.1 | 211 | Q8C351 | Q8C351 mus musculus |
| 6 | 580 | 52.0 | 246 | Q8BM17 | Q8BM17 mus musculus |
| 7 | 580 | 52.0 | 292 | Q8BVU2 | Q8BVU2 mus musculus |
| 8 | 185 | 16.0 | 1290 | Q9W6E1 | Q9W6E1 gallus gall |
| 9 | 178.5 | 16.0 | 1456 | Q61830 | Q61830 mus musculus |
| 10 | 177.5 | 15.9 | 1348 | Q25199 | Q25199 hydra atten |
| 11 | 176.5 | 15.8 | 1479 | Q9Y5P9 | Q9Y5P9 homo sapien |
| 12 | 176.5 | 15.8 | 1479 | Q9UBG0 | Q9UBG0 homo sapien |
| 13 | 170.5 | 15.3 | 1479 | Q64449 | Q64449 mus musculus |
| 14 | 169 | 15.2 | 217 | Q8C4F8 | Q8C4F8 mus musculus |
| 15 | 165.5 | 14.8 | 742 | Q8K4Q8 | Q8K4Q8 mus musculus |
| 16 | 165.5 | 14.8 | 742 | Q8C979 | Q8C979 mus musculus |

ALIGNMENTS

RESULT 1

Q8TAY8

ID Q8TAY8 PRELIMINARY; PRT; 374 AA.

AC Q8TAY8;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Similar to unnamed protein product.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Strausberg R.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC025407; AAH25407.1; -

DR InterPro; IPR001304; Lectin_C.

DR Pfam; PF00059; lectin_c; 1.

DR SMART; SM00034; CLECT; 1.

DR PROSITE; PS0041; C_TYPE_LECTIN_2; 1.

SQ SEQUENCE 374 AA; 42312 MW; FC214E8BC9E578D9 CRC64;

Query Match 100.0%; Score 1115; DB 4; Length 374;

Best Local Similarity 100.0%; Pred. No. 4.5e-99;

Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLSGQPVCRGGTQRPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSISEDEOKL 60

Db 22 ATGRLSGQPVCRGGTQRPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSISEDEOKL 81

QY 61 IEKFTENLLPSDGDWFGLRRREEKOSNSTACQDLYAWTDGSIQPRNMYVDSPSCSEV 120

Db 82 IEKFTENLLPSDGDWFGLRRREEKOSNSTACQDLYAWTDGSIQPRNMYVDSPSCSEV 141

QY 121 CVVMYHQPAPAGIGGPFYFQWNNDRCKNNFICKYSDKPAVPSPREAGETELTPV 180

Db 142 CVVMYHQPAPAGIGGPFYFQWNNDRCKNNFICKYSDKPAVPSPREAGETELTPV 201

17 161.5 14.5 134 5 Q9XYX3
18 159.5 14.3 742 11 Q8VIF6
19 158.5 14.2 719 6 O62623
20 158 14.2 142 11 Q8CJ86
21 158 14.2 142 11 Q8BHK7
22 158 14.2 295 11 Q91ZM4
23 158 14.2 311 11 Q9DBV4
24 158 14.2 325 11 Q91ZX0
25 157 14.1 158 13 Q90W17
26 156 14.0 323 11 Q8CJ91
27 156 14.0 339 6 Q95244
28 155 13.9 293 11 Q8EGZ0
29 155 13.9 323 11 Q8CJ94
30 155 13.9 323 11 Q8CJ93
31 155 13.9 323 11 Q8CJ88
32 155 13.9 1152 13 Q90WM2
33 154 13.8 322 11 Q8CJ89
34 154 13.8 323 11 Q8CJ92
35 152.5 13.7 485 6 Q95LG3
36 152 13.6 158 13 Q90W16
37 151.5 13.6 652 4 Q81XK1
38 150.5 13.5 399 6 Q8HY12
39 150 13.5 197 6 Q28008
40 149.5 13.4 158 11 Q8JZX6
41 149.5 13.4 459 5 Q22136
42 149 13.4 158 13 Q90W18
43 148.5 13.3 163 13 Q8AXR8
44 148.5 13.3 195 5 Q27340
45 148.5 13.3 330 4 Q9NT67

Q9xyx3 hydra magni
Q8vif6 mus musculus
O62623 bos taurus
Q8cj86 mus musculus
Q8bhk7 mus musculus
Q91zm4 mus musculus
Q9dbv4 mus musculus
Q91zx0 mus musculus
Q90w17 bungarus fa
Q8cj91 mus musculus
Q95244 sus scrofa
Q8bgz0 mus musculus
Q8cj94 mus musculus
Q8cj93 mus musculus
Q8cj88 mus musculus
Q90wm2 xenopus lae
Q8cj89 mus musculus
Q8cj92 mus musculus
Q95lg3 odocoileus
Q90w16 bungarus mu
Q81xk1 homo sapien
Q8hy12 hyllobates l
Q28008 bos taurus
Q8jzx6 mus musculus
Q22136 caenorhabdi
Q90w18 bungarus fa
Q8axr8 anguilla ja
Q27340 megabalanus
Q9nt67 homo sapien

| | |
|---------------------------|--|
| RA | Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S., |
| RA | Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R., |
| RA | Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., |
| RA | Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., |
| RA | Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M., |
| RA | Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., |
| RA | Kawakami B., Suzuki Y., Sugano S., Negahari K., Masuhara Y., Nagai K., |
| RA | Isoigai T.; |
| RT | "NEDO human cDNA sequencing project"; |
| RL | Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases. |
| DR | ENBL; AK055654; BAB70978.1; |
| DR | InterPro; IPR001304; Lectin_C. |
| DR | Pfam; PF00059; lectin_c; 1. |
| DR | SMART; SM00034; CLECT; 1. |
| DR | PROSITE; PS50041; C_TYPE_LECTIN_2; 1. |
| KW | Hypothetical protein. |
| SQ | SEQUENCE 374 AA; 42310 MW; CBF74E676E23BA16 CRC64; |
| | |
| Query Match | 99.6%; Score 1111; DB 4; Length 374; |
| Best Local Similarity | 99.5%; Pred. No. 1.le-98; |
| Matches 205; Conservative | 0; Mismatches 1; Indels 0; Gaps 0; |
| | |
| QY | 1 ATGRLLSGQPVCRCGTQRPCYKVIFYFHDTSRLNFEEAKACRRDGGOLVSISEDEOKL 60 |
| Db | 22 ATGRLLSGQPVCRCGTQRPCYKVIFYFHDTSRLNFEEAKACRRDGGOLVSISEDEOKL 81 |
| | |
| QY | 61 IEKFIENLLPSGDWFIGLRREEKQSNTACQDLAYMTDGSISQFRNWYVDSPSCGSEV 120 |
| Db | 82 IEKFIENLLPSGDWFIGLRREEKQSNTACQDLAYMTDGSISQFRNWYVDSPSCGSEV 141 |
| | |
| QY | 121 CVVMYHQSPAPAGIGPYMFQNNDRCNMKNFFICKYSDEKPAPVPSRAGEETELTPV 180 |
| Db | 142 CVVMYHQSPAPAGIGPYMFQNNDRCNMKNFFICKYSDEKPAPVPSRAGEETELTPV 201 |
| | |
| QY | 181 LPREETQEEDAKTKFKESREAAALNLAY 206 |
| Db | 202 LPREETQEEDTKTKFKESREAAALNLAY 227 |
| | |
| RESULT 4 | |
| Q9Z209 | |
| ID | Q9Z209 PRELIMINARY; PRT; 374 AA. |
| AC | Q9Z209; |
| DT | 01-MAY-1999 (TrEMBLrel. 10, Created) |
| DT | 01-MAY-1999 (TrEMBLrel. 10, Last sequence update) |
| DT | 01-MAR-2003 (TrEMBLrel. 23, Last annotation update) |
| DE | Laylin. |
| OS | Cricetus griseus (Chinese hamster). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; |
| OC | Cricetulus. |
| OX | NCBI_TaxID=10029; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RA | Borowsky M.L., Hynes R.O.; |
| RT | "Laylin, a novel talin-binding transmembrane protein homologous with |
| RL | C-type lectins, is localized in membrane ruffles."; |
| JL | J. Cell Biol. 143:0-0(1998). |
| DR | ENBL; AF093673; AAC68695.1; -. |
| DR | HSSP; P22897; IEGG. |
| DR | InterPro; IPR001304; Lectin_C. |
| DR | Pfam; PF00059; lectin_c; 1. |
| DR | SMART; SM00034; CLECT; 1. |
| DR | PROSITE; PS50041; C_TYPE_LECTIN_2; 1. |
| SQ | SEQUENCE 374 AA; 42435 MW; 298A8BA24FB04E1C CRC64; |
| | |
| Query Match | 84.8%; Score 945.5; DB 11; Length 374; |
| Best Local Similarity | 85.9%; Pred. No. 9.4e-83; |
| Matches 176; Conservative | 9; Mismatches 19; Indels 1; Gaps 1; |
| | |
| QY | 3 GRLLSGQPVCRGGTQRPCYKVIFYFHDTSRLNFEEAKACRRDGGOLVSISEDEOKLJE 62 |
| Db | 24 GRLLSGQLVCRGGTRRPCYKVIFYFHDAFORLNFEAKACRRDGGOLVSTETEDORLJE 83 |

QY 63 KFIENLPSDGFWIGLRRREEKQSNSTACQDLYAWTDGSIQFRNRYVDEPSCGSEVCV 122
 Db 84 KFIENLPSDGFWIGLRRREEKQSNSTACQDLYAWTDGSIQFRNRYVDEPSCGSEVCV 143
 QY 123 VMYHOPSAPAGIGGPPYMFQWDDRCNMKNFICKYSDEKPA-VPSREAEGETELTTPVL 181
 Db 144 VMYHOPSAPAGIGGPPYMFQWDDRCNMKNFICKYADEKPSITPSIRPGGEATEPPTPVL 203
 QY 182 PEETOEDDAKTKFESREAAALNAY 206
 Db 204 PEETOEDDAKTKFESREAAALNAY 228
 RESULT 5
 Q8C351 PRELIMINARY; PRT; 211 AA.
 AC Q8C351;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE LAYILIN homolog (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Lung;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RL Nature 420:563-573(2002)."
 DR EMBL; AK086930; BAC39765.1; --
 FT NON_TER 211 211
 SQ SEQUENCE 211 AA; 23697 MW; AD9870B5957DD5AE CRC64;
 Query Match 72.1%; Score 803.5; DB 11; Length 211;
 Best Local Similarity 79.8%; Pred. No. 2.2e-69;
 Matches 150; Conservative 9; Mismatches 20; Indels 9; Gaps 2;
 QY 3 GRLLS-----GQVCRGCTGTRPCYKVIYFHDTSRRLNFEAEKACRRDGGQGLVSI 54
 Db 24 GRLLSASDLDPGQGLVCRGCTGTRPCYKVIYFHDTSRRLNFEAEKACRRDGGQGLVSI 83
 QY 55 EDSOKLIEFIENLPSDGFWIGLRRREEKQSNSTACQDLYAWTDGSIQFRNRYVDEP 114
 Db 84 EDSOKLIEFIENLPSDGFWIGLRRREEKQSNSTACQDLYAWTDGSIQFRNRYVDEP 143
 QY 115 SCGSEVCVMYHOPSAPAGIGGPPYMFQWDDRCNMKNFICKYSDEKPA-VPSREAEGET 173
 Db 144 SCGSEVCVMYHOPSAPAGIGGPPYMFQWDDRCNMKNFICKYHDDKPSITPSIRPGGEA 203
 QY 174 TELTTPVL 181
 Db 204 TEPATPLL 211
 RESULT 6
 Q8BM17 PRELIMINARY; PRT; 246 AA.
 AC Q8BM17;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE C-type lectin protein MT75 homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Forelimb;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RL Nature 420:563-573(2002)."
 DR EMBL; AK031063; BAC27234.1; --
 SQ SEQUENCE 246 AA; 27394 MW; E872660C58267752 CRC64;
 Query Match 52.0%; Score 580; DB 11; Length 246;
 Best Local Similarity 57.9%; Pred. No. 8.9e-48;
 Matches 110; Conservative 26; Mismatches 40; Indels 14; Gaps 5;
 QY 4 RLLSGQPVCRGCTGTRPCYKVIYFHDTSRRLNFEAEKACRRDGGQGLVSI 63
 Db 23 RVVSGQKVCFAVDVKHPCYKWAYFHELSRSVFOEARLACESEGGVLLSLENAEOKLIES 82
 QY 64 FIENLLP-----SDGDFWIGLRRREEKQSNSTACQDLYAWTDGSIQFRNRYVDEPSCGS 118
 Db 83 MLQNLTKPGTGISDGFWIGLRRSGDGT--SGACPDLYQWSDGSSSQFRNRYTDEPSCGS 141
 QY 119 EVCVVMYHOPSAPAGIGGPPYMFQWDDRCNMKNFICKYSDE-KPAVPSREAEGETELT 177
 Db 142 EKCVMYHQPATNPGLGGPYLYQWDDRCNMKNYICKYEPIHPTPEA-----EKPVL 196
 QY 178 TPVLPEETOE 187
 Db 197 NQ--PEETHE 204
 RESULT 7
 Q8BVU2 PRELIMINARY; PRT; 292 AA.
 AC Q8BVU2;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE C-type lectin protein MT75 homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RL Nature 420:563-573(2002)."
 DR EMBL; AK076523; BAC36378.1; --
 SQ SEQUENCE 292 AA; 32502 MW; 73E631C0714D54E2 CRC64;
 Query Match 52.0%; Score 580; DB 11; Length 292;
 Best Local Similarity 57.9%; Pred. No. 1.1e-47;
 Matches 110; Conservative 26; Mismatches 40; Indels 14; Gaps 5;
 QY 4 RLLSGQPVCRGCTGTRPCYKVIYFHDTSRRLNFEAEKACRRDGGQGLVSI 63
 Db 23 RVVSGQKVCFAVDVKHPCYKWAYFHELSRSVFOEARLACESEGGVLLSLENAEOKLIES 82
 QY 64 FIENLLP-----SDGDFWIGLRRREEKQSNSTACQDLYAWTDGSIQFRNRYVDEPSCGS 118
 Db 83 MLQNLTKPGTGISDGFWIGLRRSGDGT--SGACPDLYQWSDGSSSQFRNRYTDEPSCGS 141
 QY 119 EVCVVMYHOPSAPAGIGGPPYMFQWDDRCNMKNFICKYSDE-KPAVPSREAEGETELT 177
 Db 142 EKCVMYHQPATNPGLGGPYLYQWDDRCNMKNYICKYEPIHPTPEA-----EKPVL 196
 QY 178 TPVLPEETOE 187

Db 197 NQ--PRETHE 204

RESULT 8

Q9W6E1 PRELIMINARY; PRT; 1290 AA.

AC Q9W6E1 AC Q9W6E1

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Neurocan core protein.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OC NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20309833; PubMed=10851024;

RT Li H., Leung T.C., Hoffman S., Balsamo J., Lillian J.;

RT "Coordinate Regulation of Cadherin and Integrin Function by the

RT Chondroitin Sulfate Proteoglycan Neurocan.";

RL J. Cell Biol. 149:1275-1288(2000).

DR EMBL; AF116856; AAD24546.2; -

DR HSP; P08709; Ibf9.

DR InterPro; IPR000152; Asx_hydroxyl.

DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR001881; EGF_Ca.

DR InterPro; IPR001438; EGF_II.

DR InterPro; IPR006209; EGF_Like.

DR InterPro; IPR003599; Ig.

DR InterPro; IPR007110; Ig-Like.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR001304; Lectin_C.

DR InterPro; IPR000538; Link.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00008; EGF_2.

DR Pfam; PF00047; Ig; 1.

DR Pfam; PF00059; lectin_c; 1.

DR Pfam; PF00084; sushi; 1.

DR Pfam; PF00193; Xlink; 2.

DR PRINTS; PR00010; EGFLOD.

DR PRINTS; PR01265; LINKMODULE.

DR ProDom; PD000918; Link; 2.

DR SMART; SM00032; CCP; 1.

DR SMART; SM00034; CLECT; 1.

DR SMART; SM00179; EGF_Ca; 1.

DR SMART; SM00409; IG_1.

DR SMART; SM00445; LINK; 2.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.

DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.

DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.

DR PROSITE; PS00022; EGF_1; 2.

DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS01187; EGF_Ca; 1.

DR PROSITE; PS00835; IG_LIKE; 1.

DR PROSITE; PS01241; LINK; 2.

KW EGF-like domain.

SEQUENCE 1290 AA; 182BD86D0E40BE78 CRC64;

Query Match 16.6%; Score 185; DB 13; Length 1290;

Best Local Similarity 34.7%; Pred. No. 7e-09;

Matches 50; Conservative 18; Mismatches 42; Indels 34; Gaps 8;

QY 17 QRPCYVIFHDTSRRLNFEAKEACRRDGGQGVLSIESEDEQKLEKFIENLLPSDGDWF 76

Db 1064 QGHYR-YF---SRRSWEDRRCRRAGHLTSLHSEHGFINSF-----GHNTW 1112

QY 77 IGLRREEKQSNSTACQDLYAWTDGSIQFRNYYVDEPS---CGSEVCVMY-HOPSAPA 132

Db 1113 IGLNDRIVEQD-----FQWTDNTGLQYENWRENQPDNFFAGDCDVLVSHE----- 1159

QY 133 GIGGPFYFQWDDRCNMKNFICK 156

Db 1160 -IG-----KWDVPCYNLPYICK 1177

RESULT 9

Q61830 PRELIMINARY; PRT; 1456 AA.

AC Q61830 AC Q61830

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Macrophage mannose receptor precursor.

GN MRC1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=C57BL/6;

RX MEDLINE=93043353; PubMed=1421407;

RA Harris N., Rits M., Chang G., Ezekowitz R.B.;

RT "Characterization of the murine macrophage mannose receptor.";

RL Blood 80:2363-2373(1992).

RN [2]

RP SEQUENCE FROM N.A.

RX STRAIN=C57BL/6;

RA Super M.;

RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z11974; CAA78028.1; -

DR HSP; P22897; IEGG.

DR MGD; MGI:97142; Mrcl.

DR InterPro; IPR002353; AntifreezeII.

DR InterPro; IPR000562; FN_Type_II.

DR InterPro; IPR001304; Lectin_C.

DR InterPro; IPR000772; Ricin_B_lectin.

DR Pfam; PF00040; fn2; 1.

DR Pfam; PF00059; lectin_c; 8.

DR Pfam; PF00652; Ricin_B_lectin; 2.

DR PRINTS; PR00356; ANTIFREEZEII.

DR PRINTS; PR00013; FNYPEII.

DR ProDom; PD000995; FN_Type_II; 1.

DR SMART; SM00034; CLECT; 8.

DR SMART; SM00059; FN2; 1.

DR SMART; SM00458; RICIN; 1.

DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 6.

DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 8.

DR PROSITE; PS00023; FIBONECTIN_2; 1.

DR PROSITE; PS00231; RICIN_B_LLECTIN; 1.

KW Receptor; Signal.

FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 19 1456 MACROPHAGE MANNOSE RECEPTOR.

SEQUENCE 1456 AA; 165065 MW; 4EBD3F1B8619A594 CRC64;

Query Match 16.0%; Score 178.5; DB 11; Length 1456;

Best Local Similarity 25.9%; Pred. No. 3.4e-08;

Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;

QY 21 YKVIYFHDTSRRLNFEAKEACRRDGGQGVLSIESEDEQKLEKFIENLLPSDGDWF 80

Db 807 YKDYQYVFSKEKETMDNARRFCCKNPGDLATIKSEKKFLWKYI-NKNGQSPYFIGNL 865

QY 81 RREEKQSNSTACQDLYAWTDGSIQFRNYYVDEPSCGS---EVCVMYHOPSAPAGIGGY 138

Db 866 ISMDK-----FIMWDSKVDVAVATGEPNFANDENCVTMY-----TNSGF----- 908

QY 139 MFWNDRCNMKNFICK---YSDEKPAVPVSREAGEETEELTPVLPESTQE----- 187

Db 909 ---WVDINGYPNFICQRHNSINATAMP-----TTPTPGCKRGWHLKYNK 954

QY 188 -----EDAKKTFKESREALNL 204

```
Db 955 CFKIFGFANBEKKSQDARQACKL 979
RESULT 10
Q25199 PRELIMINARY; PRT; 1348 AA.
AC Q25199;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Tyrosine kinase receptor.
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6087;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Irvine;
RX MEDLINE=20209407; PubMed=10744720;
RA Reidling J.C., Miller M.A., Steele R.E.;
RT "Sweet tooth, a Novel Receptor Protein-tyrosine Kinase with C-type
Lectin-like Extracellular Domains."
RL J. Biol. Chem. 275:10323-10330(2000).
DR EMBL; L22612; AAA29218.2; -
DR HSP; P11362; IFGK
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR003990; Pancratiss_ac.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00059; lectin_c; 4.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR01504; PNCREATITAP.
DR ProDom; PD000001; Prot_kinase; 2.
DR SMART; SM00034; CLECT; 4.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 2.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Kinase; Receptor; Transferase; Tyrosine-protein Kinase.
KW SEQUENCE 1348 AA; 156916 MW; 77D2122093227FEF CRC64;
SQ
Query Match 15.9%; Score 177.5; DB 5; Length 1348;
Best Local Similarity 28.0%; Pred. No. 3.9e-08;
Matches 47; Conservative 32; Mismatches 46; Indels 43; Gaps 8;
QY 6 LSGQPVC--RGGTQRPC-----YKVIYFHDTSRRLNFEAEACRRDGGOLVSISEDE 57
Db 412 LSHRFICKVKRATNEYCAEGWTSYRIYCYFIYIEFDWFKSFSSCONIGNLLSIENQEE 471
QY 58 QKUIEKPIENLLPSDGD-FWIGLRR-----REEKOSNSTACQDLYAWTDGSIQSFRNMY 110
Db 472 ----NRFIENDLIKNDKYWIGLKNWYDKKKNK-----FEWSDNTVTQFFNWI 518
QY 111 VDEP--SCGSEVCMVYHPSAPAGIGPWFQNDRCNMKNPFCK 156
Db 519 THQPDNNNGIESCVENYN-----GWSDECKVLNGFICK 553
RESULT 11
QY5P9 PRELIMINARY; PRT; 1479 AA.
AC QY5P9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Endocytic receptor Endo180.
GN Endo180.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
```

```
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20148849; PubMed=10683150;
RA Sheikh H., Yarwood H., Ashworth A., Isacke C.;
RT "Endo180, an endocytic recycling glycoprotein related to the
macrophage mannose receptor is expressed on fibroblasts, endothelial
cells and macrophages and functions as a lectin receptor.";
RL J. Cell Sci. 113:1021-1032(2000).
DR EMBL; AF134838; AAD30280.1; -
DR HSP; P02751; 2FN2.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000566; Lipoclin_cytFABP.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin_c; 8.
DR PRINTS; PR00013; FNTYPEII.
DR ProDom; PD000995; FN_Type_II; 1.
DR SMART; SM00034; CLECT; 8.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 3.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 8.
DR PROSITE; PS00023; FIBONECTIN_2; 1.
DR PROSITE; PS00213; LIPOCALIN; 1.
DR PROSITE; PS0231; RICIN_B_LECTIN; 1.
DR Receptor.
KW SEQUENCE 1479 AA; 166669 MW; 9F4BAF355F036FCE CRC64;
SQ
Query Match 15.8%; Score 176.5; DB 4; Length 1479;
Best Local Similarity 30.4%; Pred. No. 5.4e-08;
Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;
QY 7 SGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAEACRRDGGOLVSISEDEQKLIKXIE 66
Db 385 SWQPF-----QGHCYRL-----QAEKRSWOESKACLRGGDLVSIHSMAELEFITKQIK 434
QY 67 NLLPSDGDFFWIGLRRREEKOSNSTACQDLYAWTDGSIQSFRNMYVDEPS---CGSEVCW 123
Db 435 QEVE---ELWIGL-----NDLKLQMFWSGDSILVSTHWHFPFEPNNFRDLSLEDCVT 483
QY 124 MYHPSAPAGIGPWFQNDRCNMKNPFICKYKDEKPAVPSREAE 171
Db 484 IW---GPEG-----RWNDSPCNQSLPSICKKAGQLSQCAAEEDHG 520
RESULT 12
QYUBG0 PRELIMINARY; PRT; 1479 AA.
AC QYUBG0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Urokinase receptor-associated protein UPARAP.
GN KIAA0709.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Behrendt N., Jensen O.N., Engelholm L.H., Mortz E., Mann M., Dano K.;
RT "A urokinase receptor-associated protein with specific collagen-
binding properties.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
The complete sequences of 100 new cDNA clones from brain which can
```

```

code for large proteins in vitro."
RL DNA Res. 5:169-176(1998).
DR EMBL; AF107292; AAF14192.1; -.
DR EMBL; AB014609; BAA31684.1; -.
DR HSSP; P02751; 2FN2.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR001304; Lectin C.
DR InterPro; IPR000566; Lipocin_cytFABP.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin c; 8.
DR PRINTS; PR00013; FNTYPEII.
DR ProDom; PD000995; FN_Type_II; 1.
DR SMART; SM00034; CLECT; 8.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 3.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 8.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS00213; LIPOCALIN; 1.
DR PROSITE; PS00231; RICIN_B_LECTIN; 1.
DR Kinase; Receptor.
KW SEQUENCE 1479 AA; 166654 MW; C7583EA78E2792D1 CRC64;

Query Match 15.8%; Score 176.5; DB 4; Length 1479;
Best Local Similarity 30.4%; Pred. No. 5.4e-08;
Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;

QY 7 SQGVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGGLVSIESDEQKLEKIE 66
Db 385 SWQPF-----QHCYRL-----QAEKRSWQESKACLRGGGLVSIHSMAELEFITKQIK 434

QY 67 NLLPSDGGDFWIGLRREERKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCV 123
Db 435 QEVE---ELWIGL-----NDLKLQNFWSDCGLSVFTHWHPFEPNFRDLSLEDCVT 483

QY 124 MYHQPSAPAGIGGPPYMFQWDDRCNMKNFICKYDEKPAVPSRAEG 171
Db 484 IW----GPEG-----RWNDSPCQNSLPSICKYCKAQGLSQGAEEHDHG 520

RESULT 13
Q64449 PRELIMINARY; PRT; 1479 AA.
AC Q64449;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Lectin lambda.
GN MRC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=96355501; PubMed=8702911;
RA Wu K., Yuan J., Lasky L.A.;
RT "Characterization of a novel member of the macrophage mannose receptor
RT type C lectin family."
RL J. Biol. Chem. 271:21323-21330(1996).
DR EMBL; U56734; AAC52729.1; -.
DR HSSP; P02751; 2FN2.
DR MGD; MGI:107818; Mrc2.
DR InterPro; IPR002353; AntifreezeII.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR001304; Lectin C.
DR InterPro; IPR000566; Lipocin_cytFABP.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin c; 8.
DR PRINTS; PR00356; ANTIFREEZEII.

```

```

DR PRINTS; PR00013; FNTYPEII.
DR ProDom; PD000995; FN_Type_II; 1.
DR SMART; SM00034; CLECT; 8.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 3.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 8.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS00213; LIPOCALIN; 1.
DR PROSITE; PS00231; RICIN_B_LECTIN; 1.
DR SEQUENCE 1479 AA; 167112 MW; 62D456E10B9B48C1 CRC64;

Query Match 15.3%; Score 170.5; DB 11; Length 1479;
Best Local Similarity 31.4%; Pred. No. 2.1e-07;
Matches 48; Conservative 22; Mismatches 48; Indels 35; Gaps 7;

QY 7 SQGVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGGLVSIESDEQKLEKIE 66
Db 384 SWQPF-----QHCYRL-----QAEKRSWQESKACLRGGGLVSIHSMAELEFITKQIK 433

QY 67 NLLPSDGGDFWIGLRREERKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCV 123
Db 434 QEVE---ELWIGL-----NDLKLQNFWSDCGLSVFTHWHPFEPNFRDLSLEDCVT 482

QY 124 MYHQPSAPAGIGGPPYMFQWDDRCNMKNFICK 156
Db 483 IW----GPEG-----RWNDSPCQNSLPSICK 504

RESULT 14
Q8C4F8 PRELIMINARY; PRT; 217 AA.
AC Q8C4F8;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Chondroitin sulfate proteoglycan 3 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK082298; BAC38459.1; -.
FT NON_TER
SQ SEQUENCE 217 AA; 25949 MW; 5F1A6A57B05B76D6 CRC64;

Query Match 15.2%; Score 169; DB 11; Length 217;
Best Local Similarity 31.1%; Pred. No. 2.9e-08;
Matches 42; Conservative 16; Mismatches 47; Indels 30; Gaps 5;

QY 25 YFHTSRLNFEAEKACRRDGGGLVSIESDEQKLEKFIENLLPSDGFWIGLRREE 84
Db 3 YF---AHRRAWEADRCRRRAGHLTSVHSPEEHKFINSF-----GHENSWIGLNDRTV 53

QY 85 KQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCVVMYHQPSAPAGIGGPPYMFQ 141
Db 54 ERD-----FQWTDNTGLQYENRWREKOPNFFAGGEDCVVMVAHESG-----R 95

QY 142 WNDRCNMKNFICK 156
Db 96 WNDVPCNVLPSVCK 110

RESULT 15

```


Search completed: December 22, 2003, 16:13:31
Job time : 24.2017 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:04:13 ; Search time 52.9807 Seconds
(without alignments)
1120.478 Million cell updates/sec

Title: US-09-887-855-2

Perfect score: 2000

Sequence: 1 MRPTALQAVLLAVLLVGLR.....PDQMGRSKESGWENEIYGY 374

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq 19Jun03.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 2000 | 100.0 | 374 | 21 | AAV93948 |
| 2 | 2000 | 100.0 | 374 | 22 | AA803651 |
| 3 | 2000 | 100.0 | 374 | 23 | AB990203 |
| 4 | 1990 | 99.5 | 374 | 21 | AAV91490 |
| 5 | 1986 | 99.3 | 382 | 20 | AAV13367 |
| 6 | 1986 | 99.3 | 382 | 22 | AAU29033 |
| 7 | 1986 | 99.3 | 382 | 22 | AA802035 |
| 8 | 1986 | 99.3 | 382 | 24 | ABU69645 |
| 9 | 1986 | 99.3 | 382 | 24 | ABU71121 |

| | | | | | | |
|----|------|------|-----|----|----------|--------------------|
| 10 | 1986 | 99.3 | 382 | 24 | ABU71468 | Human PRO polypept |
| 11 | 1986 | 99.3 | 382 | 24 | ABU71914 | Human secreted/tra |
| 12 | 1986 | 99.3 | 382 | 24 | ABU65578 | Human secreted/tra |
| 13 | 1986 | 99.3 | 382 | 24 | ABU65911 | Novel human secret |
| 14 | 1986 | 99.3 | 382 | 24 | ABU67368 | Human secreted pro |
| 15 | 1986 | 99.3 | 382 | 24 | ABU67415 | Human secreted/tra |
| 16 | 1986 | 99.3 | 382 | 24 | ABU64522 | Human secreted/tra |
| 17 | 1986 | 99.3 | 382 | 24 | ABU65273 | Human PRO polypept |
| 18 | 1986 | 99.3 | 382 | 24 | ABU58409 | Human PRO polypept |
| 19 | 1986 | 99.3 | 382 | 24 | ABU55945 | Human secreted/tra |
| 20 | 1986 | 99.3 | 382 | 24 | ABU56940 | Human PRO polypept |
| 21 | 1986 | 99.3 | 382 | 24 | ABU54370 | Human secreted/tra |
| 22 | 1986 | 99.3 | 382 | 24 | ABU10519 | Human secreted/tra |
| 23 | 1982 | 99.1 | 387 | 22 | AAW25796 | Human protein sequ |
| 24 | 1979 | 99.0 | 374 | 23 | ABG56680 | Human novel polype |
| 25 | 645 | 32.2 | 273 | 21 | AAU18913 | A novel polypeptid |
| 26 | 645 | 32.2 | 273 | 22 | AAU12441 | Human PRO1890 poly |
| 27 | 645 | 32.2 | 273 | 22 | AAU73309 | Human C-type lecti |
| 28 | 645 | 32.2 | 273 | 22 | AAU87609 | Human PRO1890. Ho |
| 29 | 645 | 32.2 | 273 | 23 | ABG95934 | Human secreted/tra |
| 30 | 645 | 32.2 | 273 | 23 | AB855554 | Human angiogenesis |
| 31 | 645 | 32.2 | 273 | 23 | AB84948 | Human PRO1890 prot |
| 32 | 645 | 32.2 | 273 | 24 | ABU69084 | Human PRO polypept |
| 33 | 645 | 32.2 | 273 | 24 | ABU69107 | Human PRO polypept |
| 34 | 645 | 32.2 | 273 | 24 | ABU71589 | Human secreted pol |
| 35 | 645 | 32.2 | 273 | 24 | ABU72035 | Novel human secret |
| 36 | 645 | 32.2 | 273 | 24 | ABU72192 | Human PRO polypept |
| 37 | 645 | 32.2 | 273 | 24 | ABU66839 | Human PRO polypept |
| 38 | 645 | 32.2 | 273 | 24 | ABU67115 | Human secreted/tra |
| 39 | 645 | 32.2 | 273 | 24 | ABU59920 | Novel secreted and |
| 40 | 606 | 30.3 | 274 | 23 | ABP69211 | Human polypeptide |
| 41 | 576 | 28.8 | 232 | 22 | AA894192 | Human protein sequ |
| 42 | 336 | 16.8 | 81 | 21 | AAV91643 | Human secreted pro |
| 43 | 336 | 16.8 | 82 | 21 | AAV91557 | Human secreted pro |
| 44 | 336 | 16.8 | 115 | 21 | AAV91642 | Human secreted pro |
| 45 | 275 | 13.8 | 102 | 22 | ABG51278 | Human liver peptid |

ALIGNMENTS

| | | |
|----------|-----------------|--|
| RESULT 1 | AAV93948 | AAV93948 standard; Protein; 374 AA. |
| ID | AAV93948 | |
| XX | AC | AAV93948; |
| XX | 03-OCT-2000 | (first entry) |
| DT | XX | |
| DE | XX | Amino acid sequence of a lectin ss3939 polypeptide. |
| XX | XX | Human; lectin ss3939; chromosome 11; gene therapy. |
| XX | OS | Homo sapiens. |
| XX | XX | |
| FH | Key | Location/Qualifiers |
| FT | Peptide | 1...21 |
| FT | Domain | /note= "signal peptide" |
| FT | Domain | 22..227 |
| FT | Domain | /note= "extracellular coding region" |
| FT | Domain | 228..248 |
| FT | Domain | /note= "predicted transmembrane domain" |
| FT | Domain | 249..374 |
| FT | Domain | /note= "predicted cytoplasmic or intracellular domain" |
| XX | XX | |
| PN | WO200039296-A1. | |
| XX | XX | |
| PD | 06-JUL-2000. | |
| XX | XX | |
| PF | 22-DEC-1999; | 99WO-US30523. |
| XX | XX | |
| XX | 23-DEC-1998; | 98US-0113820. |
| XX | XX | |

```

PA (IMMV ) IMMUNEX CORP.
XX
PI Anderson DA;
XX
XX WPI; 2000-452394/39.
DR N-PSDB; AAA57382.
XX
PT ss3939 nucleic acids, polypeptides and antibodies, useful for
PT identifying human chromosome 11 and diseases associated with it -
XX
XX Claim 12; Page 8; 73pp; English.
XX
XX The present sequence represents a human lectin ss3939 polypeptide. The
XX polynucleotide sequence is a source of probes, which may be used
XX to identify nucleic acids encoding ss3939 proteins, to identify human
XX chromosome number 11, to map genes on human chromosome number 11, to
XX identify diseases associated with chromosome 11, as single-stranded
XX sense or antisense oligonucleotides to inhibit expression of
XX polypeptides encoded by the ss3939 gene, and for gene therapy. The
XX ss3939 polypeptides may be useful for developing treatments for
XX diseases (none specified) associated with defective or insufficient
XX amounts of the polypeptides. The antibodies may be useful for
XX detecting the presence of ss3939 polypeptides.
XX
SQ Sequence 374 AA;
Query Match 100.0%; Score 2000; DB 21; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.1e-184;
Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLSQGPVCRGQTQPCYKVIYFHDTSRLNFEAK 60
Db |||||||
QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLSQGPVCRGQTQPCYKVIYFHDTSRLNFEAK 60
Db |||||||
QY 61 EACRRDGGOLVSIIESEDEOKLIEKFIENLLPSDGFWMIGLRREBKQSNSTACQDIYAWT 120
Db |||||||
QY 61 EACRRDGGOLVSIIESEDEOKLIEKFIENLLPSDGFWMIGLRREBKQSNSTACQDIYAWT 120
Db |||||||
QY 121 DGSISQFRNWWYDFPSCSEVCVMYHOPSPAGIGGPMYFQWDDRCNMKNFKCKYSD 180
Db |||||||
QY 121 DGSISQFRNWWYDFPSCSEVCVMYHOPSPAGIGGPMYFQWDDRCNMKNFKCKYSD 180
Db |||||||
QY 181 EKPAPVSREAGEETELTPVLPEETQEDAKKTFKESREAAALNAYILIPISPLLLIV 240
Db |||||||
QY 181 EKPAPVSREAGEETELTPVLPEETQEDAKKTFKESREAAALNAYILIPISPLLLIV 240
Db |||||||
QY 241 VTTVVCVWVICRKRREQDPSTKQHTIWPSPHQNSPDLEVYVIRKQSEADLAETRP 300
Db |||||||
QY 241 VTTVVCVWVICRKRREQDPSTKQHTIWPSPHQNSPDLEVYVIRKQSEADLAETRP 300
Db |||||||
QY 301 DLKNISFRVCSGEATPDDMSCDYDNMNVNPSSEGFVTLVSVESGFVTNDIYFSPDQMR 360
Db |||||||
QY 301 DLKNISFRVCSGEATPDDMSCDYDNMNVNPSSEGFVTLVSVESGFVTNDIYFSPDQMR 360
Db |||||||
QY 361 SKESGWVNEIYGY 374
Db |||||||
QY 361 SKESGWVNEIYGY 374
Db |||||||
RESULT 2
AAE03651
ID AAE03651 standard; Protein; 374 AA.
XX
XX AAE03651;
XX
XX 06-AUG-2001 (first entry)
XX
XX Human extracellular matrix and cell adhesion molecule-15 (XMAP-15).
XX
XX Human; extracellular matrix and cell adhesion molecule; XMAP;
XX gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
XX Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
XX sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma;

```

```

KW inflammatory disorder; acquired immune deficiency syndrome; AIDS;
KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
KW infection; cell proliferative disorder; actinic keratosis; myeloma;
KW arteriooclerosis; nontropic; anticonvulsant; antithyroid; nephrotropic;
KW neuroprotective; dermatological.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..24
XX Protein 25..374
XX /label= Signal_peptide
XX /note= "Mature human extracellular matrix and cell
XX adhesion molecule (XMAP)"
XX 46..63
XX /note= "C-type lectin domain"
XX Domain 163..176
XX /note= "C-type lectin domain"
XX Domain 224..247
XX /note= "Transmembrane motif"
XX Domain 328..348
XX /note= "Transmembrane motif"
XX
XX WO200142285-A2.
XX
XX 14-JUN-2001.
XX
XX 05-DEC-2000; 2000WO-US32990.
XX
XX 10-DEC-1999; 99US-0172852.
XX 16-DEC-1999; 99US-0172354.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
XX Baughn MR, Lu DM, Shan P, Au-Young J;
XX WPI; 2001-381632/40.
XX N-PSDB; AAD08059.
XX
XX New human extracellular matrix and cell adhesion molecules and
XX polynucleotide sequences encoding them, useful for diagnosis,
XX prevention, treatment of genetic, autoimmune and cell proliferative
XX disorders
XX
XX Claim 1; Page 108-109; 135pp; English.
XX
XX The present sequence is a human extracellular matrix and cell
XX adhesion molecule (XMAP). The XMAP is used for screening a compound for
XX effectiveness as an agonist or antagonist of XMAP. The identified agonist
XX or antagonist are used for treating a disease or condition associated
XX with decreased or increased expression of functional XMAP. The
XX polynucleotides encoding XMAP are useful in somatic or germline gene
XX therapy to correct a genetic deficiency, to express a conditionally
XX lethal gene product and to express a protein which affords protection
XX against intracellular parasites and also for diagnosis of disorders
XX associated with expression of XMAP. They are also used for generating
XX hybridisation probes useful in mapping the naturally occurring genomic
XX sequences and to create knock in humanised animals (pigs) or transgenic
XX animals (mice or rats) to model human diseases. Oligonucleotide or longer
XX fragments derived from the polynucleotide sequences may be used as
XX elements on a microarray. Antibodies which specifically bind XMAP may be
XX used for the diagnosis of disorders associated with the expression of
XX XMAP, or in assays to monitor patients being treated with XMAP. Diseases
XX diagnosed, prevented or treated include genetic disorders such as
XX adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
XX disease, myotonic dystrophy, sickle cell anaemia, thalassaemia,
XX autoimmune/inflammatory disorders such as acquired immune deficiency
XX syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
XX atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,

```

CC glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
 CC osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
 CC bacterial, fungal, parasitic, protozoal and helminthic infections and
 CC cell proliferative disorders such as actinic keratosis, arteriosclerosis
 CC and cancer including breast, bladder, bone marrow, brain and uterus
 CC cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.
 XX
 SQ Sequence 374 AA;
 Query Match 100.0%; Score 2000; DB 22; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.1e-184;
 Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRGGTORPCYKVIYFHDTSRRLNFEAK 60
 Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRGGTORPCYKVIYFHDTSRRLNFEAK 60
 Qy 61 EACRRDGGQLVSI ESEDEQKLI EKFIE NLLPSDGF WIGLR RREEKQSNSTACQDLYAWT 120
 Db 61 EACRRDGGQLVSI ESEDEQKLI EKFIE NLLPSDGF WIGLR RREEKQSNSTACQDLYAWT 120
 Qy 121 DGSISQFNWVVDPSGSEVCVVMYHQPAPAGIGGYPMFQWDDRCNMKNFICKYSD 180
 Db 121 DGSISQFNWVVDPSGSEVCVVMYHQPAPAGIGGYPMFQWDDRCNMKNFICKYSD 180
 Qy 181 EKPAVPSREAGEETELTPVLPEETQEDAKKTFKESREALNLAYILIPSIPLLLLV 240
 Db 181 EKPAVPSREAGEETELTPVLPEETQEDAKKTFKESREALNLAYILIPSIPLLLLV 240
 Qy 241 VTTVVCWVWICRKRREOPDSTKKQHTIWPSPHQGNSPDLEVYVIRKQSEADLAETRP 300
 Db 241 VTTVVCWVWICRKRREOPDSTKKQHTIWPSPHQGNSPDLEVYVIRKQSEADLAETRP 300
 Qy 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTDIYFSPDQNGR 360
 Db 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTDIYFSPDQNGR 360
 Qy 361 SKESGWENEIYGY 374
 Db 361 SKESGWENEIYGY 374
 RESULT 3
 ID ABB90203 standard; Protein; 374 AA.
 XX
 AC ABB90203;
 XX
 DT 24-MAY-2002 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2579.
 XX
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 PN W0200190304-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 18-MAY-2001; 2001WO-US16450.
 XX
 PR 19-MAY-2000; 2000US-205515P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CE, Rosen CA;
 XX
 DR WPI; 2002-122018/16.

N-PSDB; ABL90612.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -

Claim 11; SEQ ID NO 2579; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB9040-ABB90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 374 AA;

Query Match 100.0%; Score 2000; DB 23; Length 374;
 Best Local Similarity 100.0%; Pred. No. 3.1e-184;
 Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRGGTORPCYKVIYFHDTSRRLNFEAK 60
 Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRGGTORPCYKVIYFHDTSRRLNFEAK 60
 Qy 61 EACRRDGGQLVSI ESEDEQKLI EKFIE NLLPSDGF WIGLR RREEKQSNSTACQDLYAWT 120
 Db 61 EACRRDGGQLVSI ESEDEQKLI EKFIE NLLPSDGF WIGLR RREEKQSNSTACQDLYAWT 120
 Qy 121 DGSISQFNWVVDPSGSEVCVVMYHQPAPAGIGGYPMFQWDDRCNMKNFICKYSD 180
 Db 121 DGSISQFNWVVDPSGSEVCVVMYHQPAPAGIGGYPMFQWDDRCNMKNFICKYSD 180
 Qy 181 EKPAVPSREAGEETELTPVLPEETQEDAKKTFKESREALNLAYILIPSIPLLLLV 240
 Db 181 EKPAVPSREAGEETELTPVLPEETQEDAKKTFKESREALNLAYILIPSIPLLLLV 240
 Qy 241 VTTVVCWVWICRKRREOPDSTKKQHTIWPSPHQGNSPDLEVYVIRKQSEADLAETRP 300
 Db 241 VTTVVCWVWICRKRREOPDSTKKQHTIWPSPHQGNSPDLEVYVIRKQSEADLAETRP 300
 Qy 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTDIYFSPDQNGR 360
 Db 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTDIYFSPDQNGR 360
 Qy 361 SKESGWENEIYGY 374
 Db 361 SKESGWENEIYGY 374

RESULT 4

AAY91490

ID AAY91490 standard; Protein; 374 AA.

XX

AC AAY91490;

XX

DT 29-JUN-2000 (first entry)

XX

DE Human secreted protein sequence encoded by gene 40 SEQ ID NO:163.

XX

KW Human; secreted protein; diagnosis; cytostatic; neuroprotective; immunosuppressive;
 KW antiHIV; antiinflammatory; nontropic; neuroprotective; antiallergic;
 KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
 KW antiporiatic; cardiant; gene therapy; cancer; neurological disorder;
 KW immune disease; inflammation; blood disorder; tumour.

OS Homo sapiens.

XX WO200006698-A1.

XX 10-FEB-2000.

XX 29-JUL-1999; 99WO-US17130.

XX 30-JUL-1998; 98US-0094657.

PR 05-AUG-1998; 98US-0095486.

PR 06-AUG-1998; 98US-0095454.

PR 12-AUG-1998; 98US-0095455.

XX 98US-0096319.

PA (HUMA-) HUMAN GENOME SCI INC.

XX Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;

PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;

PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;

XX WPI; 2000-195282/17.

XX N-PSDB; AAA26385.

XX New isolated human genes and the secreted polypeptides they encode,

PT useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders

XX Claim 11; Page 483-484; 634pp; English.

XX The polynucleotide sequences given in AAA26346 to AAA26458 encode the

CC human secreted proteins given in AAY91451 to AAY91691. The human secreted

CC proteins can have activities based on the tissues and cells they are

CC expressed in. Examples of the activities are: cytostatic; neuroprotective;

CC immunosuppressive; antiHIV; antiinflammatory; nontropic; neuroprotective;

CC antiarthritic; osteopathic; antiarthritic; antibacterial; antidiabetic;

CC antiasthma; antiporiatic; and cardiant. The polynucleotides and their

CC corresponding secreted proteins are useful for preventing, treating or

CC ameliorating medical conditions, e.g. by protein or gene therapy. Also

CC pathological conditions can be diagnosed by determining the amount of the

CC proteins in a sample or by determining the presence of mutations in the

CC polynucleotides. Specific uses are described for each of the

CC in, and include developing products for the diagnosis or treatment of

CC cancer, tumours, neurodegenerative disorders, developmental abnormalities

CC and foetal deficiencies, blood disorders, diseases of the immune system,

CC autoimmune diseases, hepatic and renal disease, inflammation,

CC allergies, Alzheimer's and behavioural disorders, schizophrenia,

CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,

CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,

CC cardiovascular disorders, reproductive disorders, gastrointestinal

CC disorders, respiratory disorders and metabolic disorders. The proteins

CC or polynucleotides can also be used as food additives or preservatives.

CC The proteins are also useful for identifying their binding partners.

CC AAA26337 to AAA26345 and AAY91450 are sequences used in the

CC exemplification of the present invention.

XX Sequence 374 AA;

SQ

Query Match 99.5%; Score 1990; DB 21; Length 374;

Best Local Similarity 99.5%; Pred. No. 2.9e-183;

Matches 372; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

1 MRPGTALQAVLAVLVGRLAATGRLLSGQVPCRCGTQPCVKYVIFHDTSRRLNFEAK 60

1 MRPGTALQAVLAVLVGRLAATGRLLSGQVPCRCGTQPCVKYVIFHDTSRRLNFEAK 60

61 EACRRDGGQLVSIESDEQKLIEXFIENLLPSDGFWIGLRRREKQSNSTACQDLAYMT 120

Db 61 EACRRDGGQLVSIESDEQKLIEXFIENLLPSDGFWIGLRRREKQSNSTACQDLAYMT 120
 QY 121 DGSISQFRNMYVDEPSCGSEVCVVMYHQPSAPAGIGPYNFQWDDRCNNKNNFICKYSD 180
 Db 121 DGSISQFRNMYVDEPSCGSEVCVVMYHQPSAPAGIGPYNFQWDDRCNNKNNFICKYSD 180
 QY 181 EKPAVPSREAGSEETELTPVLPEETOEDAKKTFKESREAAALNAYILIPSLPLLLLV 240
 Db 181 EKPAVPSREAGSEETELTPVLPEETOEDAKKTFKESREAAALNAYILIPSLPLLLLV 240
 QY 241 VTTVWCWVICRKRREQDPSTKQHTIWPSPHQNSPDLEVYNVIRKQSEADLSTRP 300
 Db 241 VTTVWCWVICRKRREQDPSTKQHTIWPSPHQNSPDLEVYNVIRKQSEADLSTRP 300
 QY 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFGVTNDIYFSPDQMG 360
 Db 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFGVTNDIYFSPDQMG 360
 QY 361 SKESGWVENEIYGY 374
 Db 361 SKESGWVENEIYGY 374

RESULT 5

AAV13367

ID AAY13367 standard; Protein; 382 AA.

XX AAY13367;

XX 25-JUN-1999 (first entry)

XX Amino acid sequence of protein PRO234.

XX Secreted protein; transmembrane protein; human; enterocolitis;

XX Zollinger-Ellison syndrome; gastrointestinal ulceration;

XX congenital microvillus atrophy; skin disease; cell growth;

XX abnormal keratinocyte differentiation; psoriasis; epithelial cancer;

XX Parkinson's disease; Alzheimer's disease; ALS; neuropathy;

XX fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;

XX anti-thrombotic; wound healing; tissue repair.

XX Homo sapiens.

XX WO9914328-A2.

XX 25-MAR-1999.

XX 16-SEP-1998; 98WO-US19330.

XX 25-NOV-1997; 97US-0066840.

XX 17-SEP-1997; 97US-0059113.

XX 17-SEP-1997; 97US-0059115.

XX 17-SEP-1997; 97US-0059117.

XX 17-SEP-1997; 97US-0059119.

XX 17-SEP-1997; 97US-0059121.

XX 17-SEP-1997; 97US-0059122.

XX 18-SEP-1997; 97US-0059263.

XX 18-SEP-1997; 97US-0059266.

XX 15-OCT-1997; 97US-0062125.

XX 17-OCT-1997; 97US-0062285.

XX 17-OCT-1997; 97US-0062287.

XX 21-OCT-1997; 97US-0063486.

XX 24-OCT-1997; 97US-0062814.

XX 24-OCT-1997; 97US-0062816.

XX 24-OCT-1997; 97US-0063045.

XX 24-OCT-1997; 97US-0063120.

XX 24-OCT-1997; 97US-0063121.

XX 24-OCT-1997; 97US-0063127.

XX 24-OCT-1997; 97US-0063128.

XX 27-OCT-1997; 97US-0063329.

XX 27-OCT-1997; 97US-0063327.

PR 28-OCT-1997; 97US-0063541.
 PR 28-OCT-1997; 97US-0063542.
 PR 28-OCT-1997; 97US-0063543.
 PR 28-OCT-1997; 97US-0063544.
 PR 28-OCT-1997; 97US-0063549.
 PR 28-OCT-1997; 97US-0063550.
 PR 28-OCT-1997; 97US-0063564.
 PR 29-OCT-1997; 97US-0063435.
 PR 29-OCT-1997; 97US-0063704.
 PR 29-OCT-1997; 97US-0063732.
 PR 29-OCT-1997; 97US-0063738.
 PR 29-OCT-1997; 97US-0063734.
 PR 29-OCT-1997; 97US-0064215.
 PR 29-OCT-1997; 97US-0063735.
 PR 31-OCT-1997; 97US-0063870.
 PR 31-OCT-1997; 97US-0064103.
 PR 03-NOV-1997; 97US-0064248.
 PR 07-NOV-1997; 97US-0064809.
 PR 12-NOV-1997; 97US-0065186.
 PR 17-NOV-1997; 97US-0065846.
 PR 18-NOV-1997; 97US-0065693.
 PR 21-NOV-1997; 97US-0066120.
 PR 21-NOV-1997; 97US-0066364.
 PR 24-NOV-1997; 97US-0066772.
 PR 24-NOV-1997; 97US-0066466.
 PR 24-NOV-1997; 97US-0066770.
 PR 24-NOV-1997; 97US-0066511.
 PR 24-NOV-1997; 97US-0066453.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 XX
 DR WPI; 1999-229533/19.
 DR N-PSDB; AAX52238.
 XX
 XX New isolated human genes and polypeptides used in, e.g. treatment of
 PT Gastrointestinal ulceration
 XX
 XX Claim 12; Fig 50; 320pp; English.
 XX
 CC AAY13344-403 represent secreted and transmembrane human proteins.
 CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to
 CC growth or survival of nerve cells including Parkinson's disease,
 CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
 CC for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair; PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.
 XX
 SQ Sequence 382 AA;
 Query Match 99.3%; Score 1986; DB 20; Length 382;
 Best Local Similarity 97.9%; Pred. No. 7.2e-183;
 Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQPVCRGTQPCVKIYFHDTSR 52
 DB 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLDLRGGQPVCRGTQPCVKIYFHDTSR 60
 QY 53 RLNFEEAKEACRRDGGQLVSI ESEDEQKLI EKFIENLLPSDGD F WIGLRREEKQSNSTA 112

Db 61 RLNFEEAKEACRRDGGQLVSI ESEDEQKLI EKFIENLLPSDGD F WIGLRREEKQSNSTA 120
 QY 113 QDLYAWTDGSI S QFRNMYVD E PSCGSEV CVVMYHQPSAPAGIGPYPFWNDRCNMKN 172
 Db 121 QDLYAWTDGSI S QFRNMYVD E PSCGSEV CVVMYHQPSAPAGIGPYPFWNDRCNMKN 180
 QY 173 NFICKYSD E KPAPVSR E AEG E TETLTPVLPEETO E EDAKKTFFKESREALNLA YLILPS 232
 Db 181 NFICKYSD E KPAPVSR E AEG E TETLTPVLPEETO E EDAKKTFFKESREALNLA YLILPS 240
 QY 233 I P L L L L V V T V V C W W I C R K R K R E Q P D P S T K K Q H T I W P S P H Q G N S P D L E V V N V I R K Q S E 292
 Db 241 I P L L L L V V T V V C W W I C R K R K R E Q P D P S T K K Q H T I W P S P H Q G N S P D L E V V N V I R K Q S E 300
 QY 293 A D L A E T R P D L K N I S F R V C S G E A T P D D M S C D Y D N M A V N P S E G F V T L V S V E S G F V T N D I Y E 352
 Db 301 A D L A E T R P D L K N I S F R V C S G E A T P D D M S C D Y D N M A V N P S E G F V T L V S V E S G F V T N D I Y E 360
 QY 353 F S P D Q M G R S K E S G W V E N E I Y G Y 374
 Db 361 F S P D Q M G R S K E S G W V E N E I Y G Y 382
 RESULT 6
 AAU29033
 ID AAU29033 standard; Protein; 382 AA.
 XX
 AC AAU29033;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human PRO polypeptide sequence #10.
 XX
 KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 FN WO200168848-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06520.
 XX
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 03-MAR-2000; 2000US-187202P.
 PR 06-MAR-2000; 2000US-186968P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 14-MAR-2000; 2000US-189328P.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 21-MAR-2000; 2000US-190828P.
 PR 21-MAR-2000; 2000US-191007P.
 PR 21-MAR-2000; 2000US-191048P.
 PR 21-MAR-2000; 2000US-191314P.
 PR 28-MAR-2000; 2000US-192655P.
 PR 29-MAR-2000; 2000US-193032P.
 PR 29-MAR-2000; 2000US-193053P.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 04-APR-2000; 2000US-194449P.
 PR 04-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 11-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.

PR 25-APR-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 22-AUG-2000; 2000US-0644848.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI: 2001-602746/68.
 DR N-PSDB; AAS45934.
 XX
 PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -
 XX
 PS Claim 11; Fig 20; 774pp; English.
 XX
 CC Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
 CC The PRO polypeptides and their associated nucleic acids can be used to
 CC detect the presence of a tumour in a mammal by comparing the level of
 CC expression of a PRO polypeptide in a test sample of cells from the animal
 CC and a control sample of normal cells, whereby a higher level of
 CC expression in the test sample indicates the presence of a tumour in the
 CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
 CC and rabbits but are preferably human. The polypeptides can be used to
 CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
 CC when contacted with it. A specific polypeptide can be used to stimulate
 CC the proliferation or differentiation of chondrocyte cells. The PRO
 CC proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.
 XX
 SQ Sequence 382 AA;
 Query Match 99.3%; Score 1986; DB 22; Length 382;
 Best Local Similarity 97.9%; Pred. No. 7.2e-183;
 Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGTQPCYKVIYFHTSR 52
 DB 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLDLRGGQPCVCRGGTQPCYKVIYFHTSR 60
 QY 53 RLNFEEAKACRRDGGQLVSISEDEOKLIEKFLENLLPSDGFWIGLRREEKQNSSTA 112
 DB 61 RLNFEEAKACRRDGGQLVSISEDEOKLIEKFLENLLPSDGFWIGLRREEKQNSSTA 120
 QY 113 QODLVAMTSGTSQPRNVYVDPSPSCGSEVVMYHQPAPAGIGGYFMQWDDRCNMKN 172
 DB 121 QODLVAMTSGTSQPRNVYVDPSPSCGSEVVMYHQPAPAGIGGYFMQWDDRCNMKN 180
 QY 173 NFICYKDEKPAVPSREAGESETLTPVLPEETOEDAKTFKESREAAALNAYILIPS 232
 DB 181 NFICYKDEKPAVPSREAGESETLTPVLPEETOEDAKTFKESREAAALNAYILIPS 240
 QY 233 IPLLILLVVTVVWVWVLCRRKRRQPPSTKQHTIWPSPHQGNSPDEVVNVIRKQSE 292
 DB 241 IPLLILLVVTVVWVWVLCRRKRRQPPSTKQHTIWPSPHQGNSPDEVVNVIRKQSE 300
 QY 293 ADLAETRPDLKNI SFRVCSGATPDMSCDYDNMAVNPSESGFVTLVSVEGFTVNDIYE 352
 DB 301 ADLAETRPDLKNI SFRVCSGATPDMSCDYDNMAVNPSESGFVTLVSVEGFTVNDIYE 360
 QY 353 FSPDQMGSKSGWVENEIYGY 374
 DB 361 FSPDQMGSKSGWVENEIYGY 382
 RESULT 7
 AAB80235
 ID AAB80235 standard; Protein; 382 AA.
 AC AAB80235;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Human PRO234 protein.
 XX
 KW Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
 KW antiparkinsonian nootropic; neuroprotective; vulnerary; cardiac;
 KW antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;
 KW antiarthritic; antiinfertility; antidiabetic; antiviral; diabetes;
 KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
 KW ischaemia; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200104311-A1.
 XX
 PD 18-JAN-2001.
 XX
 PF 22-FEB-2000; 2000WO-US04414.
 XX
 PR 07-JUL-1999; 99US-0143048.
 PR 26-JUL-1999; 99US-0145698.
 PR 28-JUL-1999; 99US-0146222.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 09-OCT-1999; 99WO-US23089.
 PR 25-NOV-1999; 99WO-US28214.
 PR 30-NOV-1999; 99WO-US28313.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 05-JAN-2000; 99WO-US00219.
 XX
 (GETH) GENENTECH INC.
 PA
 PI Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
 PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kijavini IJ;
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
 PI Williams PM, Wood WI;
 XX
 WPI: 2001-081051/09.
 DR N-PSDB; AAF72396.
 XX
 PT Sixty one nucleic acids encoding PRO polypeptides which are useful in
 PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
 PT squamous cell carcinoma) and neurodegenerative diseases (e.g.
 PT Alzheimer's disease) -
 XX
 PS Claim 1; Fig 50; 393pp; English.
 XX
 CC The present sequence is one of sixty one novel secreted and
 CC transmembrane PRO polypeptides. The PRO polypeptides are
 CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
 CC squamous cell carcinoma), gastrointestinal disorders (e.g.
 CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.
 CC endometrial bleeding, angiogenesis, ischaemias such as coronary
 CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,

| | |
|---------------------------|---|
| CC | rheumatoid arthritis, multiple sclerosis), infertility, AIDS and |
| CC | diabetes and retinal disorders such as retinitis pigmentosa. |
| CC | The PRO nucleic acids have applications in molecular biology, including |
| CC | use as hybridization probes, and in chromosome and gene mapping. |
| XX | |
| SQ | Sequence 382 AA; |
| Query Match | 99.3%; Score 1986; DB 22; Length 382; |
| Best Local Similarity | 97.9%; Pred. No. 7.2e-193; |
| Matches 374; Conservative | 0; Mismatches 0; Indels 8; Gaps 1; |
| QY | 1 MRPGTALQAVLLAVLLVGLRAAATGRLLS-----GQPVCRGGTQPCYKVIYFHDTSR 52 |
| Db | 1 MRPGTALQAVLLAVLLVGLRAAATGRLLSASDLRLGGQPVCRGGTQPCYKVIYFHDTSR 60 |
| QY | 53 RLNFEEAKEACRRDGGQLVSIEDSQKLIETKFIENLLPSDGFHWGLRRRREKQSNSTA 112 |
| Db | 61 RLNFEEAKEACRRDGGQLVSIEDSQKLIETKFIENLLPSDGFHWGLRRRREKQSNSTA 120 |
| QY | 113 CQDLYAWTSGISQFRNYYVDEPSCGSEVCVMYHOPAPAGIGGPMFOWNDRCNMKN 172 |
| Db | 121 CQDLYAWTSGISQFRNYYVDEPSCGSEVCVMYHOPAPAGIGGPMFOWNDRCNMKN 180 |
| QY | 173 NFICKYSDEKPAVPSREAGEETELTPVLPETQBEDAKKTFKESREAAALNLAYILIPS 232 |
| Db | 181 NFICKYSDEKPAVPSREAGEETELTPVLPETQBEDAKKTFKESREAAALNLAYILIPS 240 |
| QY | 233 IPLLALLVTVTVVWVWICKRKRQPDPTKQHTIWPSPHQNSPDLEVVNVIRKQSE 292 |
| Db | 241 IPLLALLVTVTVVWVWICKRKRQPDPTKQHTIWPSPHQNSPDLEVVNVIRKQSE 300 |
| QY | 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMVPSESGFVTLVSVESGFVTNDIYE 352 |
| Db | 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMVPSESGFVTLVSVESGFVTNDIYE 360 |
| QY | 353 FSPDQMGSKESGWENEIYGY 374 |
| Db | 361 FSPDQMGSKESGWENEIYGY 382 |
| RESULT 8 | |
| ID | ABU69645 |
| XX | ABU69645 standard; Protein; 382 AA. |
| AC | ABU69645; |
| XX | |
| DT | 05-JUN-2003 (first entry) |
| XX | |
| DE | Novel human secreted and transmembrane protein PRO234. |
| XX | |
| KW | Human; secreted and transmembrane protein; gene therapy; psoriasis; |
| KW | enterocolitis; gastrointestinal ulceration; skin disease; |
| KW | keratinocyte differentiation; epithelial cancer; Alzheimer's disease; |
| KW | squamous cell carcinoma; Parkinson's disease; inflammatory disease; |
| KW | amyotrophic lateral sclerosis; rheumatoid arthritis; asthma; |
| KW | multiple sclerosis; organ failure; atherosclerosis; cardiac injury; |
| KW | infertility; birth defect; premature aging; AIDS; cancer; |
| KW | diabetic complication; wound repair; tissue re-growth. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | US2003017463-A1. |
| XX | |
| PD | 23-JAN-2003. |
| XX | |
| PF | 11-JUL-2001; 2001US-0903640. |
| XX | |
| PR | 10-SEP-1998; 98WO-US18824. |
| PR | 14-SEP-1998; 98WO-US19177. |
| PR | 16-SEP-1998; 98WO-US19330. |
| PR | 17-SEP-1998; 98WO-US19437. |
| PR | 01-DEC-1998; 98WO-US25108. |
| PR | 08-SEP-1999; 99WO-US20594. |

| | |
|----|------------------------------|
| PR | 13-SEP-1999; 99WO-US20944. |
| PR | 15-SEP-1999; 99WO-US21090. |
| PR | 15-SEP-1999; 99WO-US21547. |
| PR | 05-OCT-1999; 99WO-US23089. |
| PR | 29-NOV-1999; 99WO-US28214. |
| PR | 30-NOV-1999; 99WO-US28313. |
| PR | 01-DEC-1999; 99WO-US28301. |
| PR | 02-DEC-1999; 99WO-US28564. |
| PR | 02-DEC-1999; 99WO-US28565. |
| PR | 16-DEC-1999; 99WO-US30095. |
| PR | 20-DEC-1999; 99WO-US30911. |
| PR | 20-DEC-1999; 99WO-US30999. |
| PR | 05-JAN-2000; 2000WO-US00219. |
| PR | 11-FEB-2000; 2000WO-US03565. |
| PR | 22-FEB-2000; 2000WO-US04414. |
| PR | 24-FEB-2000; 2000WO-US05004. |
| PR | 02-MAR-2000; 2000WO-US05841. |
| PR | 20-MAR-2000; 2000WO-US07377. |
| PR | 30-MAR-2000; 2000WO-US08439. |
| PR | 22-MAY-2000; 2000WO-US14042. |
| PR | 02-JUN-2000; 2000WO-US15264. |
| PR | 28-JUL-2000; 2000WO-US20710. |
| PR | 24-AUG-2000; 2000WO-US23328. |
| PR | 17-SEP-1997; 97US-059113P. |
| PR | 17-SEP-1997; 97US-059115P. |
| PR | 17-SEP-1997; 97US-059117P. |
| PR | 17-SEP-1997; 97US-059119P. |
| PR | 17-SEP-1997; 97US-059121P. |
| PR | 17-SEP-1997; 97US-059122P. |
| PR | 17-SEP-1997; 97US-059184P. |
| PR | 18-SEP-1997; 97US-059263P. |
| PR | 18-SEP-1997; 97US-059266P. |
| PR | 15-OCT-1997; 97US-062125P. |
| PR | 17-OCT-1997; 97US-062285P. |
| PR | 17-OCT-1997; 97US-062287P. |
| PR | 21-OCT-1997; 97US-063486P. |
| PR | 24-OCT-1997; 97US-062814P. |
| PR | 24-OCT-1997; 97US-062816P. |
| PR | 24-OCT-1997; 97US-063045P. |
| PR | 24-OCT-1997; 97US-063120P. |
| PR | 24-OCT-1997; 97US-063121P. |
| PR | 24-OCT-1997; 97US-063127P. |
| PR | 24-OCT-1997; 97US-063128P. |
| PR | 27-OCT-1997; 97US-063327P. |
| PR | 27-OCT-1997; 97US-063329P. |
| PR | 28-OCT-1997; 97US-063541P. |
| PR | 28-OCT-1997; 97US-063542P. |
| PR | 28-OCT-1997; 97US-063544P. |
| PR | 28-OCT-1997; 97US-063549P. |
| PR | 28-OCT-1997; 97US-063550P. |
| PR | 28-OCT-1997; 97US-063564P. |
| PR | 29-OCT-1997; 97US-063435P. |
| PR | 29-OCT-1997; 97US-063704P. |
| PR | 29-OCT-1997; 97US-063732P. |
| PR | 29-OCT-1997; 97US-063734P. |
| PR | 29-OCT-1997; 97US-063735P. |
| PR | 29-OCT-1997; 97US-063738P. |
| PR | 29-OCT-1997; 97US-064215P. |
| PR | 31-OCT-1997; 97US-063870P. |
| PR | 31-OCT-1997; 97US-064103P. |
| PR | 03-NOV-1997; 97US-064248P. |
| PR | 07-NOV-1997; 97US-064809P. |
| PR | 12-NOV-1997; 97US-065186P. |
| PR | 17-NOV-1997; 97US-065846P. |
| PR | 18-NOV-1997; 97US-065693P. |
| PR | 21-NOV-1997; 97US-066120P. |
| PR | 21-NOV-1997; 97US-066364P. |
| PR | 24-NOV-1997; 97US-066453P. |
| PR | 24-NOV-1997; 97US-066466P. |
| PR | 24-NOV-1997; 97US-066511P. |
| PR | 24-NOV-1997; 97US-066770P. |
| PR | 24-NOV-1997; 97US-066772P. |
| PR | 25-NOV-1997; 97US-066840P. |

PR 12-DEC-1997; 97US-069425P.
 PR 04-JUN-1998; 98US-088026P.
 PR 10-SEP-1998; 98US-099803P.
 PR 14-SEP-1998; 98US-100262P.
 PR 17-SEP-1998; 98US-100858P.
 PR 13-OCT-1998; 98US-104080P.
 PR 20-NOV-1998; 98US-109304P.
 PR 22-DEC-1998; 98US-113296P.
 PR 07-JUL-1999; 98US-143048P.
 PR 26-JUL-1999; 98US-145698P.
 PR 28-JUL-1999; 98US-146222P.
 PR 18-SEP-2000; 2000US-0665350.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Aashkenazi A, Botstein D, Desnovers L, Eaton DL, Ferrara N,
 PI Pilvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kijavini IJ;
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
 PI Williams PM, Wood WI;
 XX
 WIPI: 2003-341586/32.
 DR N-PSDB; ACA54920.
 XX
 PT New PRO polypeptides and nucleic acid molecules, useful in diagnosing
 PT or treating inflammatory diseases, organ failure, atherosclerosis,
 PT cardiac injury, infertility, cancer, AIDS, Alzheimer's disease or
 PT Parkinson's disease -
 XX
 PS Claim 12; Fig 50; 473pp; English.
 XX
 CC The invention describes sixty one nucleic acids encoding PRO polypeptides
 CC (secreted and transmembrane). The PRO polypeptides and nucleic acids are
 CC useful in diagnosing or treating enterocolitis, gastrointestinal
 CC ulceration, skin diseases associated with abnormal keratinocyte
 CC differentiation, e.g. psoriasis or epithelial cancers such as squamous
 CC cell carcinoma, Alzheimer's disease, Parkinson's disease, amyotrophic
 CC lateral sclerosis, inflammatory diseases, e.g. rheumatoid arthritis,
 CC asthma or multiple sclerosis, organ failure, atherosclerosis, cancer,
 CC injury, infertility, birth defects, premature aging, AIDS, cancer,
 CC diabetic complications, or mutations in general. The polypeptides are
 CC also useful for wound repair and associated therapies concerned with
 CC re-growth of tissue. The PRO polypeptides and nucleic acid molecules
 CC are also useful in gene therapy, and as molecular weight markers for
 CC protein electrophoresis purposes. The anti-PRO antibodies may be used
 CC in diagnostic assays for PRO, or for the affinity purification of PRO
 CC from recombinant cell culture or natural sources. This is the amino
 CC acid sequence of a novel human PRO polypeptide.
 XX
 SQ Sequence 382 AA;

Query Match 99.3%; Score 1986; DB 24; Length 382;
 Best Local Similarity 97.9%; Pred. No. 7.2e-183;
 Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGTQPCYKVIYFHTSR 52
 DB 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHTSR 60
 QY 53 RLNFEEAKERRDGGQLVSIIESEDEKLIKFENLLPSDGFWIGLRREKQSNSTA 112
 DB 61 RLNFEEAKERRDGGQLVSIIESEDEKLIKFENLLPSDGFWIGLRREKQSNSTA 120
 QY 113 CODLYAMTDGSIQFRNMYVDFPSCGSEVVMYHQPAPAGIGPYFMQNDRCNNKN 172
 DB 121 CODLYAMTDGSIQFRNMYVDFPSCGSEVVMYHQPAPAGIGPYFMQNDRCNNKN 180
 QY 173 NFICKYSEKPAVPSREAEGEETELTPVLPEETOEDAKTKFKESREAAALNAYILIPS 232
 DB 181 NFICKYSEKPAVPSREAEGEETELTPVLPEETOEDAKTKFKESREAAALNAYILIPS 240
 QY 233 IPLLILLVVTTVCWWMICRKRKQPDSTKKQHTIWPSPHQGNSPDLEVYVIRKQSE 292
 DB 241 IPLLILLVVTTVCWWMICRKRKQPDSTKKQHTIWPSPHQGNSPDLEVYVIRKQSE 300
 QY 293 ADLAETRPDLKNISFRVCSGEATPDMSCDYDNMAVNPSESGFVTILSVESGFVTNDIYE 352
 DB 301 ADLAETRPDLKNISFRVCSGEATPDMSCDYDNMAVNPSESGFVTILSVESGFVTNDIYE 360
 QY 353 FSPDQMGSKESGWENEIYG 374
 DB 361 FSPDQMGSKESGWENEIYG 382

RESULT 9

ABU71121

ID ABU71121 standard; Protein; 382 AA.

XX

AC ABU71121;

DT 10-JUN-2003 (first entry)

XX

DE Human PRO234 protein.

XX

KW Human; PRO; secreted; transmembrane; cytosolic; TNF-alpha; blood;
 KW tumour necrosis factor alpha release; chondrocyte cell; proliferation;
 KW differentiation; tumour; gene therapy.

XX

OS Homo sapiens.

XX

PN US2003036143-A1.

XX

PD 20-FEB-2003.

XX

PF 02-JUL-2002; 2002US-0187600.

XX

PR 16-SEP-1998; 98WO-US19330.

PR

PR 07-OCT-1998; 98WO-US21141.

PR

PR 01-DEC-1998; 98WO-US25108.

PR

PR 08-MAR-1999; 99WO-US05028.

PR

PR 14-MAY-1999; 99WO-US10733.

PR

PR 02-JUN-1999; 99WO-US12252.

PR

PR 15-SEP-1999; 99WO-US20111.

PR

PR 01-DEC-1999; 99WO-US21090.

PR

PR 02-DEC-1999; 99WO-US28301.

PR

PR 30-DEC-1999; 99WO-US31274.

PR

PR 05-JAN-2000; 2000WO-US00219.

PR

PR 18-FEB-2000; 2000WO-US04341.

PR

PR 18-FEB-2000; 2000WO-US04342.

PR

PR 22-FEB-2000; 2000WO-US04414.

PR

PR 24-FEB-2000; 2000WO-US05004.

PR

PR 01-MAR-2000; 2000WO-US05601.

PR

PR 02-MAR-2000; 2000WO-US05841.

PR

PR 15-MAR-2000; 2000WO-US06884.

PR

PR 30-MAR-2000; 2000WO-US08439.

PR

PR 17-MAY-2000; 2000WO-US13705.

PR

PR 22-MAY-2000; 2000WO-US14042.

PR

PR 30-MAY-2000; 2000WO-US14941.

PR

PR 02-JUN-2000; 2000WO-US15264.

PR

PR 28-JUL-2000; 2000WO-US20710.

PR

PR 24-AUG-2000; 2000WO-US23328.

PR

PR 08-NOV-2000; 2000WO-US30952.

PR

PR 01-DEC-2000; 2000WO-US32678.

PR

PR 20-DEC-2000; 2000WO-US34956.

PR

PR 28-FEB-2001; 2001WO-US06520.

PR

PR 01-JUN-2001; 2001WO-US17800.

PR

PR 20-JUN-2001; 2001WO-US19692.

PR

PR 29-JUN-2001; 2001WO-US21066.

PR

PR 09-JUL-2001; 2001WO-US21735.

PR

PR 29-AUG-2001; 2001WO-US27099.

PR

PR 18-SEP-1997; 97US-059263P.

PR

PR 18-SEP-1997; 97US-059266P.

PR

PR 17-OCT-1997; 97US-062250P.

PR

PR 21-OCT-1997; 97US-063486P.

PR

PR 24-OCT-1997; 97US-063120P.

PR

PR 24-OCT-1997; 97US-0631121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 28-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069425P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083435P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088266P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088455P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.

PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 01-JUL-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 02-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 01-SEP-1998; 98US-098014P.
PR 01-SEP-1998; 98US-09816P.
PR 01-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.
PR 10-SEP-1998; 98US-099754P.
PR 10-SEP-1998; 98US-099763P.
PR 10-SEP-1998; 98US-099812P.

Query Match 99.3%; Score 1986; DB 24; Length 382;
Best Local Similarity 97.9%; Pred. No. 7.2e-183;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQPVCRGQTQPCYKVIYFHDTSR 52

Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLDRGGQPVCRGGTQPCYKVIYFHDTSR 60
QY 53 RLNFEEAKEACRRDGGQLVSIIESEDEQKLIIEKFIEINLLPSDGDFFWIGLRRREKQSNSTA 112
Db 61 RLNFEEAKEACRRDGGQLVSIIESEDEQKLIIEKFIEINLLPSDGDFFWIGLRRREKQSNSTA 120
QY 113 CQDLYAWTDGSIQFRNWTVDPSGSEVCMVYHQPAPAGIGGYPYMFQWMDRCNMKN 172
Db 121 CQDLYAWTDGSIQFRNWTVDPSGSEVCMVYHQPAPAGIGGYPYMFQWMDRCNMKN 180
QY 173 NFIKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNAYILIPS 232
Db 181 NFIKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNAYILIPS 240
QY 233 IPLLILLVVTTCVWVICRKRKREQPDSTKKQHTIWPSPHQNSPDLEVNVRKQSE 292
Db 241 IPLLILLVVTTCVWVICRKRKREQPDSTKKQHTIWPSPHQNSPDLEVNVRKQSE 300
QY 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFTVNDIYE 352
Db 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFTVNDIYE 360
QY 353 FSPDQMGSKESGWVENEIYGY 374
Db 361 FSPDQMGSKESGWVENEIYGY 382

RESULT 10

ABU71468
ID ABU71468 standard; Protein; 382 AA.
AC ABU71468;
XX
DT 10-JUN-2003 (first entry)
XX
DE Human PRO polypeptide #24.
XX
KW Human; secreted and transmembrane protein; PRO polypeptide; cancer;
KW Alzheimer's disease; ischaemia; cytostatic; neurotropic; vasotropic;
KW neuroprotective.

XX Homo sapiens.

OS US2002192659-A1.

PN 19-DEC-2002.

PD 10-JUL-2001; 2001US-0902853.

PF 10-SEP-1998; 98WO-US18824.

PR 14-SEP-1998; 98WO-US19177.

PR 16-SEP-1998; 98WO-US19330.

PR 17-SEP-1998; 98WO-US19437.

PR 01-DEC-1998; 98WO-US21508.

PR 08-SEP-1999; 99WO-US20594.

PR 13-SEP-1999; 99WO-US21090.

PR 15-SEP-1999; 99WO-US21547.

PR 05-OCT-1999; 99WO-US23089.

PR 01-DEC-1999; 99WO-US28301.

PR 02-DEC-1999; 99WO-US28564.

PR 02-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 20-DEC-1999; 99WO-US30999.

PR 05-JAN-2000; 2000WO-US00219.

PR 11-FEB-2000; 2000WO-US03565.

PR 22-FEB-2000; 2000WO-US04414.

PR 28-JUL-2000; 2000WO-US20710.

PR 24-AUG-2000; 2000WO-US23328.

PR 17-SEP-1997; 97US-059113P.

PR 17-SEP-1997; 97US-059115P.

PR 17-SEP-1997; 97US-059117P.
PR 18-SEP-1997; 97US-059266P.
PR 15-OCT-1997; 97US-062125P.
PR 17-OCT-1997; 97US-062285P.
PR 17-OCT-1997; 97US-062287P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-062814P.
PR 24-OCT-1997; 97US-062816P.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi A, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
PI Filyaroff B, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kljavin LJ;
PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
PI Williams PM, Wood WI;
XX
DR WP1; 2003-361832/34.
DR N-PSDB; ACA58405.
XX
PT New isolated nucleic acid encoding a PRO polypeptide, e.g. PRO245 or
PT PRO1868, useful in molecular biology, chromosome and gene mapping, in
PT generating antisense RNA and DNA, and in gene therapy
XX
PS Claim 12; Fig 50; 474pp; English.
XX
CC The present invention relates to the isolation of novel human secreted
CC and transmembrane proteins (PRO polypeptides), and the polynucleotide
CC sequences encoding them. The polynucleotide sequences are useful in
CC molecular biology, as hybridisation probes, in chromosome and gene
CC mapping, in generating antisense RNA and DNA, and in gene therapy. The
CC polynucleotide sequences may also be used in preparing PRO polypeptides
CC by recombinant techniques, and in generating either transgenic animals
CC or knock-out animals which, in turn, are useful in the development and
CC screening of therapeutically useful reagents. The PRO polypeptides or
CC their antibodies are useful in preparing a medicament for treating a
CC condition responsive to the polypeptide or antibody, such as cancer,
CC Alzheimer's disease or ischaemia, and in various diagnostic assays.
CC ABU71445-ABU71505 represent human PRO polypeptides of the invention.
XX
SQ Sequence 382 AA;

Query Match 99.3%; Score 1986; DB 24; Length 382;
Best Local Similarity 97.9%; Pred. No. 7.2e-183;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQPVCRGGTQPCYKVIYFHDTSR 52
Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLDRGGQPVCRGGTQPCYKVIYFHDTSR 60
QY 53 RLNFEEAKEACRRDGGQLVSIIESEDEQKLIIEKFIEINLLPSDGDFFWIGLRRREKQSNSTA 112
Db 61 RLNFEEAKEACRRDGGQLVSIIESEDEQKLIIEKFIEINLLPSDGDFFWIGLRRREKQSNSTA 120
QY 113 CQDLYAWTDGSIQFRNWTVDPSGSEVCMVYHQPAPAGIGGYPYMFQWMDRCNMKN 172
Db 121 CQDLYAWTDGSIQFRNWTVDPSGSEVCMVYHQPAPAGIGGYPYMFQWMDRCNMKN 180
QY 173 NFIKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNAYILIPS 232
Db 181 NFIKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNAYILIPS 240
QY 233 IPLLILLVVTTCVWVICRKRKREQPDSTKKQHTIWPSPHQNSPDLEVNVRKQSE 292
Db 241 IPLLILLVVTTCVWVICRKRKREQPDSTKKQHTIWPSPHQNSPDLEVNVRKQSE 300
QY 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFTVNDIYE 352
Db 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFTVNDIYE 360
QY 353 FSPDQMGSKESGWVENEIYGY 374
Db 361 FSPDQMGSKESGWVENEIYGY 382

RESULT 11
 ABU71914
 ID ABU71914 standard; Protein; 382 AA.
 XX AC ABU71914;
 XX DT 12-JUN-2003 (first entry)
 XX DE Human secreted/transmembrane protein PRO234.
 XX KW Human; secreted protein; transmembrane protein; PRO;
 KW Gene therapy; chromosome identification; chromosome marker.
 XX OS Homo sapiens.
 XX PN US2003003530-A1.
 XX PD (2-JAN-2003.
 XX PF 11-JUL-2001; 2001US-0904011.
 XX PR 10-SEP-1998; 98WO-US18824.
 PR 14-SEP-1998; 98WO-US19177.
 PR 16-SEP-1998; 98WO-US19330.
 PR 17-SEP-1998; 98WO-US19437.
 PR 01-DEC-1998; 98WO-US25108.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 05-OCT-1999; 99WO-US23089.
 PR 29-NOV-1999; 99WO-US28214.
 PR 30-NOV-1999; 99WO-US28313.
 PR 01-DEC-1999; 99WO-US28301.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 02-MAR-2000; 2000WO-US05004.
 PR 20-MAR-2000; 2000WO-US05841.
 PR 30-MAR-2000; 2000WO-US07377.
 PR 22-MAY-2000; 2000WO-US08439.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 17-SEP-1997; 97US-059113P.
 PR 17-SEP-1997; 97US-059115P.
 PR 17-SEP-1997; 97US-059117P.
 PR 17-SEP-1997; 97US-059119P.
 PR 17-SEP-1997; 97US-059121P.
 PR 17-SEP-1997; 97US-059122P.
 PR 18-SEP-1997; 97US-059184P.
 PR 18-SEP-1997; 97US-059263P.
 PR 15-OCT-1997; 97US-062125P.
 PR 17-OCT-1997; 97US-062285P.
 PR 17-OCT-1997; 97US-062287P.
 PR 21-OCT-1997; 97US-063486P.
 PR 24-OCT-1997; 97US-062814P.
 PR 24-OCT-1997; 97US-062816P.
 PR 24-OCT-1997; 97US-063045P.
 PR 24-OCT-1997; 97US-063120P.
 PR 24-OCT-1997; 97US-063121P.
 PR 24-OCT-1997; 97US-063127P.
 PR 24-OCT-1997; 97US-063128P.
 PR 27-OCT-1997; 97US-063327P.

PR 27-OCT-1997; 97US-063329P.
 PR 28-OCT-1997; 97US-063541P.
 PR 28-OCT-1997; 97US-063542P.
 PR 28-OCT-1997; 97US-063544P.
 PR 28-OCT-1997; 97US-063549P.
 PR 28-OCT-1997; 97US-063550P.
 PR 28-OCT-1997; 97US-063564P.
 PR 29-OCT-1997; 97US-063435P.
 PR 29-OCT-1997; 97US-063704P.
 PR 29-OCT-1997; 97US-063732P.
 PR 29-OCT-1997; 97US-063734P.
 PR 29-OCT-1997; 97US-063735P.
 PR 29-OCT-1997; 97US-063738P.
 PR 31-OCT-1997; 97US-064215P.
 PR 31-OCT-1997; 97US-063870P.
 PR 03-NOV-1997; 97US-064103P.
 PR 07-NOV-1997; 97US-064248P.
 PR 07-NOV-1997; 97US-064809P.
 PR 12-NOV-1997; 97US-065186P.
 PR 17-NOV-1997; 97US-065846P.
 PR 18-NOV-1997; 97US-065693P.
 PR 21-NOV-1997; 97US-066120P.
 PR 24-NOV-1997; 97US-066364P.
 PR 24-NOV-1997; 97US-066453P.
 PR 24-NOV-1997; 97US-066466P.
 PR 24-NOV-1997; 97US-066511P.
 PR 24-NOV-1997; 97US-066770P.
 PR 18-SEP-2000; 97US-066772P.
 XX
 (GETH) GENENTECH INC.

Ashkenazi A, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
 Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kijavini IG;
 Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
 Williams PM, Wood WI;
 WPI; 2003-329602/31.
 N-PSDB; ACA60112.

New transmembrane polypeptides and nucleic acids encoding the
 polypeptides, useful in gene therapy, in chromosome identification, as
 chromosome markers, in generating probes and in tissue typing -
 Claim 12; Fig 50; 484pp; English.

The invention relates to an isolated nucleic acid with at least 80%
 nucleic acid sequence identity to a nucleotide sequence encoding one of
 61 secreted/transmembrane polypeptides, or PRO polypeptides or encoding a
 PRO protein extracellular domain. Also included are a vector comprising a
 the PRO nucleic acid, a host cell comprising the vector, producing a PRO
 polypeptide (by culturing the host cell for the expression of the PRO
 polypeptide, and recovering the PRO polypeptide from the cell culture),
 an isolated PRO polypeptide (having at least 80% sequence identity
 to: (a) an amino acid sequence selected from the 61 PRO proteins;
 (b) an amino acid sequence encoded by a nucleic acid molecule deposited
 with an ATCC number (detailed in the specification); or (c) an
 extracellular domain of a PRO polypeptide or to a PRO polypeptide lacking
 its associated signal peptide), a chimeric molecule comprising a PRO
 polypeptide of fused to a heterologous amino acid sequence, an anti-PRO
 antibody, detecting a PRO245 or PRO1868 in a sample suspected of
 containing the polypeptide, linking a bioactive molecule to a cell
 expressing a PRO245 or PRO1868 and modulating at least one biological
 activity of a cell expressing a PRO245 or PRO1868. Nucleic acids which
 encode PRO can be used to generate either transgenic animals or knock-out
 animals which may be used in the development and screening of
 therapeutically useful reagents. The nucleic acids may also be used in
 gene therapy, in chromosome identification, as chromosome markers, or in
 generating probes. The PRO polypeptides are useful as molecular markers
 for protein electrophoresis, and the isolated nucleic acids may be used
 for recombinantly expressing those markers. The PRO polypeptides and
 nucleic acids may also be used in tissue typing. Anti-PRO antibodies

CC are useful in diagnostic assays for PRO, and in affinity purification
CC of PRO from recombinant cell culture or natural sources. The
CC present sequence represents a PRO protein.

XX Sequence 382 AA;

Query Match 99.3%; Score 1986; DB 24; Length 382;
Best Local Similarity 97.9%; Pred. No. 7.2e-183;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 MRPGTALQAVLLAVLLVGLRAAATGRLLS-----GQVCRGQTQPCPKVIYFHDTSR 52
Db 1 MRPGTALQAVLLAVLLVGLRAAATGRLLSASDLRLGQGPVCRGQTQPCPKVIYFHDTSR 60
QY 53 RLNFEEAKEACRRDGGQVLSIEDEQKLEKFIENLLPSDGFQWGLRREKQNSTA 112
Db 61 RLNFEEAKEACRRDGGQVLSIEDEQKLEKFIENLLPSDGFQWGLRREKQNSTA 120
QY 113 QDLYAWTDGSIQFRNWWYDEPSCGSEVVCVMYHQPSPAGIGGYPMFQWDDRCNMKN 172
Db 121 QDLYAWTDGSIQFRNWWYDEPSCGSEVVCVMYHQPSPAGIGGYPMFQWDDRCNMKN 180
QY 173 NFICKYSDEKPAVPSREAGEEETELTPVLPSTOEDAKTKFKESREAAINLAYILIPS 232
Db 181 NFICKYSDEKPAVPSREAGEEETELTPVLPSTOEDAKTKFKESREAAINLAYILIPS 240
QY 233 IPLLILLVTVTVVVCWMIWCRKRKREQDPSTKKOHTIWPSPHOGNPDLEVVNVIRKQSE 292
Db 241 IPLLILLVTVTVVVCWMIWCRKRKREQDPSTKKOHTIWPSPHOGNPDLEVVNVIRKQSE 300
QY 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNNAVNPSESGFVTLVSVEGFTVNDIYE 352
Db 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNNAVNPSESGFVTLVSVEGFTVNDIYE 360
QY 353 FSPDQGRSKESGWENEIYGY 374
Db 361 FSPDQGRSKESGWENEIYGY 382

RESULT 12

ABU65578

ID ABU65578 standard; Protein; 382 AA.

XX AC ABU65578;

XX DT 19-MAY-2003 (first entry)

XX DE Human secreted/transmembrane protein, SEQ ID 20.

XX KW Human; PRO; secreted protein; transmembrane protein;
KW cytosolic; antiarthritic; osteopathic; adrenal tumour; lung tumour;
KW colon tumour; breast tumour; prostate tumour; rectal tumour;
KW cervical tumour; liver tumour; TNF-alpha release; arthritis;
KW tumour necrosis factor alpha; chondrocyte cell; bone disorder;
KW cartilage disorder; sports injury.

XX OS Homo sapiens.

XX PN US2003036156-A1.

XX PD 20-FEB-2003.

XX PF 02-JUL-2002; 2002US-0188767.

XX PP 16-SEP-1998; 98WO-US19330.

PR 07-OCT-1998; 98WO-US21141.

PR 01-DEC-1998; 98WO-US25108.

PR 08-MAR-1999; 99WO-US05028.

PR 14-MAY-1999; 99WO-US10733.

PR 02-JUN-1999; 99WO-US12252.

PR 01-SEP-1999; 99WO-US20111.

PR 15-SEP-1999; 99WO-US21090.

PR 01-DEC-1999; 99WO-US28301.

PR 02-DEC-1999; 99WO-US28551.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04342.
PR 24-FEB-2000; 2000WO-US04414.
PR 01-MAR-2000; 2000WO-US05004.
PR 02-MAR-2000; 2000WO-US05601.
PR 15-MAR-2000; 2000WO-US05841.
PR 30-MAR-2000; 2000WO-US06884.
PR 17-MAY-2000; 2000WO-US08439.
PR 22-MAY-2000; 2000WO-US13705.
PR 30-MAY-2000; 2000WO-US14042.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 29-AUG-2001; 2001WO-US27099.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 29-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 11-DEC-1997; 97US-066772P.
PR 12-DEC-1997; 97US-069335P.
PR 17-DEC-1997; 97US-069425P.
PR 18-DEC-1997; 97US-069870P.
PR 10-MAR-1998; 97US-068017P.
PR 11-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 20-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078986P.
PR 27-MAR-1998; 98US-078939P.
PR 31-MAR-1998; 98US-079664P.
PR 31-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 01-APR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081818P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083495P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.

```
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089651P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 02-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089651P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 02-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
Query Match 99.3%; Score 1986; DB 24; Length 382;
Beat Local Similarity 97.9%; Pred. No. 7.2e-183;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
Qy 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----CQPVCRGGTORPCYKVIYFPHDTSR 52
Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLRLGGQPVCRGGTORPCYKVIYFPHDTSR 60
Qy 53 RLNFEEAKEACRRDGGQLVSIESTDEOKLIEKFTIENLLPSDGFWIGLRRREEKQSNSTA 112
Db 61 RLNFEEAKEACRRDGGQLVSIESTDEOKLIEKFTIENLLPSDGFWIGLRRREEKQSNSTA 120
Qy 113 CQDLYAMTDGSIQFRNMYVDPSGSEVCVVMYTHQPSAPAGICGPFYMFQWDDRCNMKN 172
Db 121 CQDLYAMTDGSIQFRNMYVDPSGSEVCVVMYTHQPSAPAGICGPFYMFQWDDRCNMKN 180
Qy 173 NFICKYSDEKPAVPSREAGEETELTPVLPEETQBEDAKKTFKESREAAALNAYILIPS 232
Db 181 NFICKYSDEKPAVPSREAGEETELTPVLPEETQBEDAKKTFKESREAAALNAYILIPS 240
Qy 233 IPLILLVTVVVCWWVICRKRREQDPSTKKQHTIWPSPHQNSPDLEVYVIRKQSE 292
Db 241 IPLILLVTVVVCWWVICRKRREQDPSTKKQHTIWPSPHQNSPDLEVYVIRKQSE 300
Qy 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFTNDIYE 352
Db 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFTNDIYE 360
Qy 353 FSPDQGRSKESGWENEIYGY 374
Db 361 FSPDQGRSKESGWENEIYGY 382
RESULT 13
ABU65911
ID ABU65911 standard; Protein; 382 AA.
XX
XX ABU65911;
XX
XX 20-MAY-2003 (first entry)
XX
XX Novel human secreted and transmembrane protein PRO234.
XX Human; secreted protein; transmembrane protein; cytostatic;
KW gene therapy; TNF-Agonist-Alpha; chondrocyte stimulator; tumour;
KW adrenal tumour; lung tumour; colon tumour; breast tumour;
KW prostate tumour; rectal tumour; cervical tumour; liver tumour.
XX
XX Homo sapiens.
XX
```

| | | | | | |
|------------------|--------------|-----------------|---------------|--------------|---------------|
| US2003036157-A1. | 02-JUL-2002; | 2002US-0188769. | 98US-0810499; | 08-APR-1998; | 98US-0810499; |
| 20-FEB-2003. | 16-SEP-1998; | 98WO-US19330. | 98US-081070P; | 08-APR-1998; | 98US-081070P; |
| | 07-OCT-1998; | 98WO-US21141. | 98US-081195P; | 09-APR-1998; | 98US-081195P; |
| | 08-DEC-1998; | 98WO-US25108. | 98US-081838P; | 15-APR-1998; | 98US-081838P; |
| | 08-MAR-1999; | 99WO-US05028. | 98US-082568P; | 21-APR-1998; | 98US-082568P; |
| | 14-MAY-1999; | 99WO-US10733. | 98US-082569P; | 22-APR-1998; | 98US-082569P; |
| | 02-JUN-1999; | 99WO-US12252. | 98US-082704P; | 22-APR-1998; | 98US-082704P; |
| | 01-SEP-1999; | 99WO-US20111. | 98US-082797P; | 28-APR-1998; | 98US-082797P; |
| | 15-SEP-1999; | 99WO-US21090. | 98US-083322P; | 28-APR-1998; | 98US-083322P; |
| | 01-DEC-1999; | 99WO-US28301. | 98US-083495P; | 29-APR-1998; | 98US-083495P; |
| | 02-DEC-1999; | 99WO-US28551. | 98US-083496P; | 29-APR-1998; | 98US-083496P; |
| | 30-DEC-1999; | 99WO-US31274. | 98US-083559P; | 05-MAY-1998; | 98US-083559P; |
| | 05-JAN-2000; | 2000WO-US00219. | 98US-083436P; | 06-MAY-1998; | 98US-084414P; |
| | 18-FEB-2000; | 2000WO-US04341. | 98US-084640P; | 07-MAY-1998; | 98US-084639P; |
| | 18-FEB-2000; | 2000WO-US04342. | 98US-084643P; | 07-MAY-1998; | 98US-084643P; |
| | 22-FEB-2000; | 2000WO-US04414. | 98US-085579P; | 15-MAY-1998; | 98US-085579P; |
| | 24-FEB-2000; | 2000WO-US05004. | 98US-085580P; | 15-MAY-1998; | 98US-085580P; |
| | 01-MAR-2000; | 2000WO-US05601. | 98US-085582P; | 15-MAY-1998; | 98US-085582P; |
| | 02-MAR-2000; | 2000WO-US05841. | 98US-085700P; | 15-MAY-1998; | 98US-085700P; |
| | 15-MAR-2000; | 2000WO-US06884. | 98US-086023P; | 18-MAY-1998; | 98US-086023P; |
| | 30-MAR-2000; | 2000WO-US08439. | 98US-086392P; | 22-MAY-1998; | 98US-086392P; |
| | 22-MAY-2000; | 2000WO-US13705. | 98US-086486P; | 22-MAY-1998; | 98US-086486P; |
| | 22-MAY-2000; | 2000WO-US14042. | 98US-087098P; | 28-MAY-1998; | 98US-087098P; |
| | 02-JUN-2000; | 2000WO-US15264. | 98US-087208P; | 28-MAY-1998; | 98US-087208P; |
| | 24-JUL-2000; | 2000WO-US20710. | 98US-087609P; | 02-JUN-1998; | 98US-087609P; |
| | 24-AUG-2000; | 2000WO-US23328. | 98US-087759P; | 02-JUN-1998; | 98US-087759P; |
| | 08-NOV-2000; | 2000WO-US03952. | 98US-087827P; | 03-JUN-1998; | 98US-087827P; |
| | 01-DEC-2000; | 2000WO-US32678. | 98US-088025P; | 04-JUN-1998; | 98US-088025P; |
| | 20-DEC-2000; | 2000WO-US34956. | 98US-088028P; | 04-JUN-1998; | 98US-088028P; |
| | 28-FEB-2001; | 2001WO-US06520. | 98US-088033P; | 04-JUN-1998; | 98US-088033P; |
| | 01-JUN-2001; | 2001WO-US17800. | 98US-088326P; | 04-JUN-1998; | 98US-088326P; |
| | 20-JUN-2001; | 2001WO-US19692. | 98US-088167P; | 05-JUN-1998; | 98US-088167P; |
| | 29-JUN-2001; | 2001WO-US21066. | 98US-088202P; | 05-JUN-1998; | 98US-088202P; |
| | 09-JUL-2001; | 2001WO-US21735. | 98US-088212P; | 05-JUN-1998; | 98US-088212P; |
| | 18-SEP-1997; | 97US-059263P. | 98US-088217P; | 09-JUN-1998; | 98US-088217P; |
| | 18-SEP-1997; | 97US-059266P. | 98US-088555P; | 09-JUN-1998; | 98US-088555P; |
| | 17-OCT-1997; | 97US-062250P. | 98US-088722P; | 10-JUN-1998; | 98US-088722P; |
| | 21-OCT-1997; | 97US-063486P. | 98US-088738P; | 10-JUN-1998; | 98US-088738P; |
| | 24-OCT-1997; | 97US-063120P. | 98US-088811P; | 10-JUN-1998; | 98US-088811P; |
| | 24-OCT-1997; | 97US-063540P. | 98US-088824P; | 10-JUN-1998; | 98US-088824P; |
| | 28-OCT-1997; | 97US-063841P. | 98US-088825P. | 10-JUN-1998; | 98US-088825P. |
| | 28-OCT-1997; | 97US-063564P. | 98US-088826P. | 10-JUN-1998; | 98US-088826P. |
| | 28-OCT-1997; | 97US-063564P. | 98US-088861P; | 11-JUN-1998; | 98US-088861P; |
| | 31-OCT-1997; | 97US-063734P. | 98US-088633P; | 11-JUN-1998; | 98US-088633P; |
| | 29-OCT-1997; | 97US-063870P. | 98US-088876P; | 11-JUN-1998; | 98US-088876P; |
| | 31-OCT-1997; | 97US-064103P. | 98US-089090P; | 12-JUN-1998; | 98US-089090P; |
| | 13-NOV-1997; | 97US-065311P. | 98US-089105P; | 12-JUN-1998; | 98US-089105P; |
| | 21-NOV-1997; | 97US-066 | | | |

```
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096765P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098071P.
PR 01-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.
PR 10-SEP-1998; 98US-099754P.
PR 10-SEP-1998; 98US-099763P.

Query Match 99.38; Score 1986; DB 24; Length 382;
Best Local Similarity 97.99; Pred. No. 7.2e-183;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 MRPTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGTQPCVKVIYFHDTSR 52
Db 1 MRPTALQAVLLAVLLVGLRAATGRLLSASDLRLGGQPCVCRGGTQPCVKVIYFHDTSR 60

Qy 53 RLNFEAKEACRRDGGQGVLSIESEDEQKLIETIENLLPSDGFWIGLRREEKQSNSTA 112
Db 61 RLNFEAKEACRRDGGQGVLSIESEDEQKLIETIENLLPSDGFWIGLRREEKQSNSTA 120

Qy 113 CQDIYAWTDGSIQFRNWNVYDEPSCGSEVCVMVHQPASAGICGPFYMFQWDDRCNMKN 172
Db 121 CQDIYAWTDGSIQFRNWNVYDEPSCGSEVCVMVHQPASAGICGPFYMFQWDDRCNMKN 180

Qy 173 NFICKYSDEKPAVPSREAGEETELTPVLPETQEDDAKTKPKESREAAINLAYILIPS 232
Db 181 NFICKYSDEKPAVPSREAGEETELTPVLPETQEDDAKTKPKESREAAINLAYILIPS 240

Qy 233 IPLLLLLVTVVVCVWVICRKRREKQDPDSTKKQHTIWPSPHQNSPDLEVINVRKQSE 292
Db 241 IPLLLLLVTVVVCVWVICRKRREKQDPDSTKKQHTIWPSPHQNSPDLEVINVRKQSE 300

Qy 293 ADLAETRPDLKNLSFRVCSGEATPDDMSCDYDNMAYNPSESGFVTLVSVEGFTVNDIYE 352
Db 301 ADLAETRPDLKNLSFRVCSGEATPDDMSCDYDNMAYNPSESGFVTLVSVEGFTVNDIYE 360

Qy 353 FSPDQMGSRKESGWENEIYGY 374
Db 361 FSPDQMGSRKESGWENEIYGY 382
```

RESULT 14
ABU67368
ID ABU67368 standard; Protein; 382 AA.
XX AC ABU67368;
XX DT 29-MAY-2003 (first entry)
XX DE Human secreted protein PRO234.
XX KW Human; gene therapy; mucosal lesion; ulcer; enterocolitis; skin disease;
KW psoriasis; cancer; lung cancer; colon cancer; nerve cell disease;
KW Alzheimer's disease; Parkinson's disease; Usher syndrome; angiogenesis;
KW atrophla areata; inflammatory disease; asthma; rheumatoid arthritis;
KW ischaemia.
XX OS Homo sapiens.
XX PN US2003023054-A1.
XX PD 30-JAN-2003.
XX PF 16-JUL-2001; 2001US-0906742.
XX PR 10-SEP-1998; 98WO-US18824.
PR 14-SEP-1998; 98WO-US19177.
PR 16-SEP-1998; 98WO-US19330.
PR 17-SEP-1998; 98WO-US19437.
PR 01-DEC-1998; 98WO-US25108.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 29-NOV-1999; 99WO-US28214.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 05-JAN-2000; 2000WO-US00219.
PR 11-FEB-2000; 2000WO-US03565.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US05004.
PR 02-MAR-2000; 2000WO-US05841.
PR 20-MAR-2000; 2000WO-US07377.
PR 30-MAR-2000; 2000WO-US08439.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 17-SEP-1997; 97US-059113P.
PR 17-SEP-1997; 97US-059115P.
PR 17-SEP-1997; 97US-059117P.
PR 17-SEP-1997; 97US-059119P.
PR 17-SEP-1997; 97US-059121P.
PR 17-SEP-1997; 97US-059122P.
PR 17-SEP-1997; 97US-059184P.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 15-OCT-1997; 97US-062125P.
PR 17-OCT-1997; 97US-062285P.
PR 17-OCT-1997; 97US-062287P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-062814P.
PR 24-OCT-1997; 97US-062816P.
PR 24-OCT-1997; 97US-063045P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 24-OCT-1997; 97US-063127P.
PR 24-OCT-1997; 97US-063128P.

PR 27-OCT-1997; 97US-063327P.
 PR 27-OCT-1997; 97US-063329P.
 PR 28-OCT-1997; 97US-063541P.
 PR 28-OCT-1997; 97US-063542P.
 PR 28-OCT-1997; 97US-063544P.
 PR 28-OCT-1997; 97US-063549P.
 PR 28-OCT-1997; 97US-063550P.
 PR 28-OCT-1997; 97US-063564P.
 PR 28-OCT-1997; 97US-063435P.
 PR 29-OCT-1997; 97US-063704P.
 PR 29-OCT-1997; 97US-063732P.
 PR 29-OCT-1997; 97US-063734P.
 PR 29-OCT-1997; 97US-063735P.
 PR 29-OCT-1997; 97US-063738P.
 PR 29-OCT-1997; 97US-064215P.
 PR 31-OCT-1997; 97US-063870P.
 PR 31-OCT-1997; 97US-064103P.
 PR 03-NOV-1997; 97US-064248P.
 PR 07-NOV-1997; 97US-064809P.
 PR 12-NOV-1997; 97US-065186P.
 PR 17-NOV-1997; 97US-065846P.
 PR 18-NOV-1997; 97US-065693P.
 PR 21-NOV-1997; 97US-066120P.
 PR 21-NOV-1997; 97US-066364P.
 PR 24-NOV-1997; 97US-066453P.
 PR 24-NOV-1997; 97US-066466P.
 PR 24-NOV-1997; 97US-066511P.
 PR 24-NOV-1997; 97US-066770P.
 PR 24-NOV-1997; 97US-066772P.
 PR 25-NOV-1997; 97US-066840P.
 PR 12-DEC-1997; 97US-069425P.
 PR 04-JUN-1998; 98US-088026P.
 PR 10-SEP-1998; 98US-099803P.
 PR 14-SEP-1998; 98US-100282P.
 PR 17-SEP-1998; 98US-100858P.
 PR 13-OCT-1998; 98US-104080P.
 PR 20-NOV-1998; 98US-109304P.
 PR 22-DEC-1998; 98US-113296P.
 PR 07-JUL-1999; 99US-143048P.
 PR 26-JUL-1999; 99US-145698P.
 PR 28-JUL-1999; 99US-146222P.
 PR 18-SEP-2000; 2000US-0665350.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashtenazi A, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
 PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kljavin IJ;
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
 PI Williams PM, Wood WI;
 XX WPI: 2003-331485/31.
 DR N-PSDB; ACA05450.
 XX
 PT Sixty one isolated nucleic acids encoding a PRO polypeptide, e.g.
 PT PRO245 or PRO1868, useful in chromosome and gene mapping, in generating
 PT antisense RNA and DNA, and in treating cancer and Alzheimer's disease -
 XX
 XX Example 22; Fig 50; 481pp; English.
 XX
 CC The invention relates to sixty one nucleic acids encoding PRO
 CC polypeptides (secreted and transmembrane). The polynucleotide is useful
 CC in molecular biology, including uses as hybridisation probes, in
 CC chromosome and gene mapping, in generating antisense RNA and DNA, and in
 CC gene therapy. The polynucleotide may also be used in preparing PRO
 CC polypeptides by recombinant techniques, and in generating either
 CC transgenic animals or knock-out animals which, in turn, are useful in the
 CC development and screening of therapeutically useful reagents. The PRO
 CC polypeptide or the antibody is used in preparing a medicament for
 CC treating a condition responsive to the polypeptide or antibody, such as
 CC mucosal lesions e.g. ulcers and enterocolitis, skin disease e.g.
 CC psoriasis, cancer e.g. lung cancer and colon cancer, nerve cell disease
 CC e.g. Alzheimer's disease and Parkinson's disease, Usher syndrome.

CC atrophica areata, angiogenesis, inflammation, inflammatory disease e.g. asthma and
 CC rheumatoid arthritis, ischaemia, and in various diagnostic assays. The
 CC present sequence represents the amino acid sequence of a PRO polypeptide.
 XX
 SQ Sequence 382 AA;
 Query Match 99.3%; Score 1986; DB 24; Length 382;
 Best Local Similarity 97.9%; Pred. No. 7.2e-183;
 Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQPVCRGTQRPCKYVFFHDTSR 52
 Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLRLGQGPVCRGTQRPCKYVFFHDTSR 60
 QY 53 RLNFEEAKACRRDGGQLVSIIEDEQKLIKFIENLLPSDGFGLRRREKQSNSTA 112
 Db 61 RLNFEEAKACRRDGGQLVSIIEDEQKLIKFIENLLPSDGFGLRRREKQSNSTA 120
 QY 113 CQDLVYAWTDGSIQPRNMYVDEPSCGSEVCVVMYHQPAPAGIGGPFYMFQWDDRCNMKN 172
 Db 121 CQDLVYAWTDGSIQPRNMYVDEPSCGSEVCVVMYHQPAPAGIGGPFYMFQWDDRCNMKN 180
 QY 173 NFICKYSDEKPAVPSREAEGETELTTPVLPBETQEDAKTTPKESREAAALNLAYILIPS 232
 Db 181 NFICKYSDEKPAVPSREAEGETELTTPVLPBETQEDAKTTPKESREAAALNLAYILIPS 240
 QY 233 IPLLILLVTTVVCWWICRKRKREOPDPSTKKQHTIMFSPHQNSPDLEVNVIRKQSE 292
 Db 241 IPLLILLVTTVVCWWICRKRKREOPDPSTKKQHTIMFSPHQNSPDLEVNVIRKQSE 300
 QY 293 ADLAETRPDLKNISPRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFTVNDIYE 352
 Db 301 ADLAETRPDLKNISPRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFTVNDIYE 360
 QY 353 FSPDQMGSRKSGWVENEIYGY 374
 Db 361 FSPDQMGSRKSGWVENEIYGY 382
 RESULT 15
 ABU67415
 ID ABU67415 standard; Protein; 382 AA.
 XX
 AC ABU67415;
 XX
 DT 29-MAY-2003 (first entry)
 DE Human secreted/transmembrane protein (PRO) #10.
 XX Human; secreted and transmembrane protein; PRO; TNF-alpha;
 KW tumour necrosis factor alpha; chondrocyte cell; tumour; gene therapy;
 KW tissue typing.
 XX
 OS Homo sapiens.
 XX
 PN US2003036162-A1.
 XX
 PD 20-FEB-2003.
 XX
 PF 12-JUL-2002; 2002US-0194423.
 XX
 PR 16-SEP-1998; 98WO-US19330.
 PR 07-OCT-1998; 98WO-US21141.
 PR 01-DEC-1998; 98WO-US25108.
 PR 08-MAR-1999; 99WO-US05028.
 PR 14-MAY-1999; 99WO-US10733.
 PR 02-JUN-1999; 99WO-US12252.
 PR 01-SEP-1999; 99WO-US20111.
 PR 15-SEP-1999; 99WO-US21090.
 PR 01-DEC-1999; 99WO-US28301.
 PR 02-DEC-1999; 99WO-US28551.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.

PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 PR 28-FEB-2001; 2001WO-US05520.
 PR 01-JUN-2001; 2001WO-US17800.
 PR 20-JUN-2001; 2001WO-US19692.
 PR 29-JUN-2001; 2001WO-US21066.
 PR 09-JUL-2001; 2001WO-US21735.
 PR 29-AUG-2001; 2001WO-US27099.
 PR 26-JUN-1998; 98US-0105413.
 PR 07-OCT-1998; 98US-0168978.
 PR 06-NOV-1998; 98US-0187368.
 PR 07-DEC-1998; 98US-0202054.
 PR 03-MAR-1999; 99US-0254311.
 PR 14-MAY-1999; 99US-0311832.
 PR 14-MAY-1999; 99US-0380137.
 PR 25-AUG-1999; 99US-0380138.
 PR 25-AUG-1999; 99US-0380139.
 PR 25-AUG-1999; 99US-0380142.
 PR 18-OCT-1999; 99US-0403297.
 PR 12-NOV-1999; 99US-0423844.
 PR 22-AUG-2000; 2000US-0644848.
 PR 18-SEP-2000; 2000US-0664610.
 PR 18-SEP-2000; 2000US-0665350.
 PR 08-NOV-2000; 2000US-0709238.
 PR 20-DEC-2000; 2000US-0747259.
 PR 22-MAR-2001; 2001US-0816744.
 PR 10-MAY-2001; 2001US-0854208.
 PR 25-MAY-2001; 2001US-0866028.
 PR 05-JUN-2001; 2001US-0874503.
 PR 18-JUL-2001; 2001US-0908827.
 PR 30-JUL-2001; 2001US-0918585.
 PR 06-AUG-2001; 2001US-0924419.
 PR 13-AUG-2001; 2001US-0929404.
 PR 16-AUG-2001; 2001US-0931836.
 PR 28-AUG-2001; 2001US-0941992.
 PR 04-SEP-2001; 2001US-0946374.
 PR 15-JAN-2002; 2002US-0052586.

(GETH) GENENTECH INC.

Baker KP, Chen J, Deanoyers L, Goddard A, Godowski PJ, Gurney AL;
 Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;

WPI: 2003-332039/31.

N-PSDB; ACA05709.

XX New secreted and transmembrane PRO polypeptides and nucleic acids,
 PT useful in gene therapy, in chromosome and gene mapping, as chromosome
 PT markers, in tissue typing, and in chromosome identification -
 XX

Claim 11; Fig 20; 706pp; English.

XX The invention discloses human nucleic acids encoding secreted and
 CC transmembrane (PRO) polypeptides. Also disclosed is an antibody that
 CC specifically binds to the PRO polypeptide, a method for stimulating the
 CC release of tumour necrosis factor alpha (TNF-alpha) from human blood by
 CC contacting the blood a PRO polypeptide, a method for stimulating the

CC proliferation or differentiation of chondrocyte cells by contacting the
 CC cells with a PRO polypeptide, a method for detecting the presence of a
 CC tumour in a mammal and an oligonucleotide probe derived from any of the
 CC PRO nucleotide sequences. The nucleotide sequences are useful as probes,
 CC in chromosome and gene mapping, in generating antisense RNA and DNA, in
 CC preparing PRO polypeptides by recombinant techniques and in gene therapy
 CC (e.g. for replacement of defective gene). The PRO polypeptides are useful
 CC as molecular weight markers for protein electrophoresis purposes, for
 CC chromosome identification, as chromosome markers, as therapeutic agents,
 CC for stimulating the release of TNF-alpha from human blood, for
 CC stimulating the proliferation or differentiation of chondrocytes and
 CC detecting the presence of a tumour. The PRO polypeptides and nucleic
 CC acids may also be used diagnostically for tissue typing. The sequences
 CC presented in ABU67406-ABU67710 are the PRO polypeptides of the invention.

XX Sequence 382 AA;

Query Match 99.3%; Score 1986; DB 24; Length 382;
 Best Local Similarity 97.9%; Pred. No. 7, 2e-183;
 Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGTQRCYKVIYFHDTSR 52
 Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLRLGGQVCRGGTQRCYKVIYFHDTSR 60
 QY 53 RLNFEEAKEACRRDGGQLVSIETSEDEQKLIETFIENLLPSDGFWIGLRRREKOSNSTA 112
 Db 61 RLNFEEAKEACRRDGGQLVSIETSEDEQKLIETFIENLLPSDGFWIGLRRREKOSNSTA 120
 QY 113 QDLYAWTDGSIQFRNMYVDPEPCSGSEVCVVMYHQPAPAGIGGPFQWDDRCNNKN 172
 Db 121 QDLYAWTDGSIQFRNMYVDPEPCSGSEVCVVMYHQPAPAGIGGPFQWDDRCNNKN 180
 QY 173 NFICKYDEKPAVPSREAEGETELTTPVLPEETQEDAKKTFKESREAAALNLAIVLIPS 232
 Db 181 NFICKYDEKPAVPSREAEGETELTTPVLPEETQEDAKKTFKESREAAALNLAIVLIPS 240
 QY 233 IPLLILLVTTVCWVWVICRKRKEQDPSTKKOHTIWPSPHQGNSPDEVNVIRKOSE 292
 Db 241 IPLLILLVTTVCWVWVICRKRKEQDPSTKKOHTIWPSPHQGNSPDEVNVIRKOSE 300
 QY 293 ADLAETRPDLKNI SFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFTVNDIYE 352
 Db 301 ADLAETRPDLKNI SFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFTVNDIYE 360
 QY 353 FSPDMGRSKESGWVENEIYGY 374
 Db 361 FSPDMGRSKESGWVENEIYGY 382

Search completed: December 22, 2003, 16:10:06
 Job time : 54.9807 Secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:04:13 ; Search time 28.8986 Seconds
(without alignment)
1120.478 Million cell updates/sec

Title: US-09-887-855-2_COPY_24_227

Perfect score: 1106

Sequence: 1 GRLLSGPVCVCGGTQRPCK.....EEDAKTFKESREAAALNAY 204

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 1106 | 100.0 | 374 | 21 | AAV93948 |
| 2 | 1106 | 100.0 | 374 | 22 | AAE033651 |
| 3 | 1106 | 100.0 | 374 | 23 | ABG66680 |
| 4 | 1106 | 100.0 | 374 | 23 | ABB90203 |
| 5 | 1106 | 100.0 | 387 | 22 | AAAM25796 |
| 6 | 1106 | 99.1 | 374 | 21 | AAV91490 |
| 7 | 1092 | 98.7 | 382 | 20 | AAV13367 |
| 8 | 1092 | 98.7 | 382 | 22 | AAU29033 |
| 9 | 1092 | 98.7 | 382 | 22 | AAH80235 |

| | | | | | | |
|----|-------|------|-----|----|----------|---------------------|
| 10 | 1092 | 98.7 | 382 | 24 | ABU69645 | Novel human secret |
| 11 | 1092 | 98.7 | 382 | 24 | ABU71121 | Human PRO2334 prote |
| 12 | 1092 | 98.7 | 382 | 24 | ABU71468 | Human PRO polypept |
| 13 | 1092 | 98.7 | 382 | 24 | ABU71914 | Human secreted/tra |
| 14 | 1092 | 98.7 | 382 | 24 | ABU65578 | Human secreted/tra |
| 15 | 1092 | 98.7 | 382 | 24 | ABU65911 | Novel human secret |
| 16 | 1092 | 98.7 | 382 | 24 | ABU67368 | Human secreted pro |
| 17 | 1092 | 98.7 | 382 | 24 | ABU67415 | Human secreted/tra |
| 18 | 1092 | 98.7 | 382 | 24 | ABU64522 | Human secreted/tra |
| 19 | 1092 | 98.7 | 382 | 24 | ABU65273 | Human PRO polypept |
| 20 | 1092 | 98.7 | 382 | 24 | ABU58409 | Human PRO polypept |
| 21 | 1092 | 98.7 | 382 | 24 | ABU55945 | Human secreted/tra |
| 22 | 1092 | 98.7 | 382 | 24 | ABU56940 | Human PRO polypept |
| 23 | 1092 | 98.7 | 382 | 24 | ABU54370 | Human secreted/tra |
| 24 | 1092 | 98.7 | 382 | 24 | ABU10519 | Human secreted/tra |
| 25 | 561.5 | 50.8 | 273 | 21 | AAI18913 | A novel polypeptid |
| 26 | 561.5 | 50.8 | 273 | 22 | AAU12441 | Human PRO1890 poly |
| 27 | 561.5 | 50.8 | 273 | 22 | AAB73309 | Human C-type lecti |
| 28 | 561.5 | 50.8 | 273 | 22 | AAB7609 | Human PRO1890. Ho |
| 29 | 561.5 | 50.8 | 273 | 23 | ABG95934 | Human secreted/tra |
| 30 | 561.5 | 50.8 | 273 | 23 | ABG95534 | Human angiogenesis |
| 31 | 561.5 | 50.8 | 273 | 23 | ABB84948 | Human PRO1890 prot |
| 32 | 561.5 | 50.8 | 273 | 24 | ABU69084 | Human PRO polypept |
| 33 | 561.5 | 50.8 | 273 | 24 | ABU69107 | Human PRO polypept |
| 34 | 561.5 | 50.8 | 273 | 24 | ABU71589 | Human secreted pol |
| 35 | 561.5 | 50.8 | 273 | 24 | ABU72035 | Novel human secret |
| 36 | 561.5 | 50.8 | 273 | 24 | ABU72192 | Human PRO polypept |
| 37 | 561.5 | 50.8 | 273 | 24 | ABU66839 | Human PRO polypept |
| 38 | 561.5 | 50.8 | 273 | 24 | ABU67115 | Human secreted/tra |
| 39 | 561.5 | 50.8 | 273 | 24 | ABU59920 | Novel secreted and |
| 40 | 534.5 | 48.3 | 274 | 23 | ABG69211 | Human polypeptide |
| 41 | 504.5 | 45.6 | 222 | 22 | AA894192 | Human protein sequ |
| 42 | 275 | 24.9 | 102 | 22 | ABG51278 | Human liver peptid |
| 43 | 275 | 24.9 | 102 | 22 | ABB21788 | Protein #3787 enco |
| 44 | 275 | 24.9 | 102 | 22 | AAW29941 | Peptide #3978 enco |
| 45 | 231 | 20.9 | 81 | 21 | AAV91643 | Human secreted pro |

ALIGNMENTS

RESULT 1

AAV93948

ID AAV93948 standard; Protein; 374 AA.

XX AAV93948;

AC AAV93948;

XX 03-OCT-2000 (first entry)

DT

XX Amino acid sequence of a lectin ss3939 polypeptide.

DE Human; lectin ss3939; chromosome 11; gene therapy.

XX Homo sapiens.

OS

XX Key

XX Location/Qualifiers

FT Peptide

FT /note= "signal peptide"

FT Domain

FT /note= "extracellular coding region"

FT Domain

FT /note= "predicted transmembrane domain"

FT Domain

FT /note= "predicted cytoplasmic or intracellular domain"

FT

XX WO2000039296-A1.

PN

XX 06-JUL-2000.

PD

XX 22-DEC-1999;

PF 99WO-US30523.

XX 23-DEC-1998;

PR 98US-0113820.

XX

PA (IMMV) IMMUNEX CORP.
 XX Anderson DA;
 PI WPI: 2000-452394/39.
 DR N-PSDB; AAS7382.
 XX
 XX s9339 nucleic acids, polypeptides and antibodies, useful for
 PT identifying human chromosome 11 and diseases associated with it -
 PT
 PS Claim 12; Page 8; 73pp; English.
 XX
 XX The present sequence represents a human lectin s9339 polypeptide. The
 CC polynucleotide sequence is a source of probes, which may be used
 CC to identify nucleic acids encoding s9339 proteins, to identify human
 CC chromosome number 11, to map genes on human chromosome number 11, to
 CC identify diseases associated with chromosome 11, as single-stranded
 CC sense or antisense oligonucleotides to inhibit expression of
 CC polypeptides encoded by the s9339 gene, and for gene therapy. The
 CC s9339 polypeptides may be useful for developing treatments for
 CC diseases (none specified) associated with defective or insufficient
 CC amounts of the polypeptides. The antibodies may be useful for
 CC detecting the presence of s9339 polypeptides.
 XX
 XX Sequence 374 AA;
 SQ
 Query Match 100.0%; Score 1106; DB 21; Length 374;
 Best Local Similarity 100.0%; Pred. No. 2.5e-104;
 Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GRLLSQPVCRGGTQPCYKVIYFHTSRRLNFEBAKEACRRGGQGVLSIESDEQKLI 60
 DB 24 GRLLSQPVCRGGTQPCYKVIYFHTSRRLNFEBAKEACRRGGQGVLSIESDEQKLI 83
 QY 61 KFTENLLPSDGDPMGLRREERKQNSTACQDLYAWTDGSIQFRNWWYDESCGSEVCV 120
 DB 84 KFTENLLPSDGDPMGLRREERKQNSTACQDLYAWTDGSIQFRNWWYDESCGSEVCV 143
 QY 121 VMVQPSAPAGIGGPPYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTTPVLP 180
 DB 144 VMVQPSAPAGIGGPPYFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTTPVLP 203
 QY 181 EETQESDAKTKFESREALNLAY 204
 DB 204 EETQESDAKTKFESREALNLAY 227
 RESULT 2
 AAE03651
 ID AAE03651 standard; Protein; 374 AA.
 AC AAE03651;
 XX
 XX 06-AUG-2001 (first entry)
 DT
 XX
 DE Human extracellular matrix and cell adhesion molecule-15 (XMAD-15).
 XX
 KW Human; extracellular matrix and cell adhesion molecule; XMAD;
 KW gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
 KW Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
 KW sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma;
 KW inflammatory disorder; acquired immune deficiency syndrome; AIDS;
 KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
 KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
 KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
 KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
 KW infection; cell proliferative disorder; actinic keratosis; myeloma;
 KW arteriosclerosis; neutropenic; anticonvulsant; antithyroid; nephrotropic;
 KW neuroprotective; dermatological.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH

FT Peptide 1..24 /label= Signal_peptide
 FT PT 25..374
 FT Protein /note= "Mature human extracellular matrix and cell
 FT adhesion molecule (XMAD)"
 FT 46..63
 FT Domain /note= "C-type lectin domain"
 FT 163..176
 FT Domain /note= "C-type lectin domain"
 FT 224..247
 FT Domain /note= "Transmembrane motif"
 FT 328..348
 FT /note= "Transmembrane motif"
 XX WO200142285-A2.
 XX 14-JUN-2001.
 XX 05-DEC-2000; 2000WO-US32990.
 XX 10-DEC-1999; 99US-0172852.
 XX 16-DEC-1999; 99US-0172354.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
 PI Baughn MR, Lu DAM, Shah P, Au-Young J;
 XX WPI: 2001-381632/40.
 DR N-PSDB; AAD08059.
 XX
 XX New human extracellular matrix and cell adhesion molecules and
 PT polynucleotide sequences encoding them, useful for diagnosis,
 PT prevention, treatment of genetic, autoimmune and cell proliferative
 PT disorders
 XX
 XX Claim 1; Page 108-109; 135pp; English.
 XX
 XX The present sequence is a human extracellular matrix and cell
 CC adhesion molecule (XMAD). The XMAD is used for screening a compound for
 CC effectiveness as an agonist or antagonist of XMAD. The identified agonist
 CC or antagonist are used for treating a disease or condition associated
 CC with decreased or increased expression of functional XMAD. The
 CC polynucleotides encoding XMAD are useful in somatic or germline gene
 CC therapy to correct a genetic deficiency, to express a conditionally
 CC lethal gene product and to express a protein which affords protection
 CC against intracellular parasites and also for diagnosis of disorders
 CC associated with expression of XMAD. They are also used for generating
 CC hybridisation probes useful in mapping the naturally occurring genomic
 CC sequences and to create knock in humanised animals (pigs) or transgenic
 CC animals (mice or rats) to model human diseases. Oligonucleotide or longer
 CC fragments derived from the polynucleotide sequences may be used as
 CC elements on a microarray. Antibodies which specifically bind XMAD may be
 CC used for the diagnosis of disorders associated with the expression of
 CC XMAD, or in assays to monitor patients being treated with XMAD. Diseases
 CC diagnosed, prevented or treated include genetic disorders such as
 CC adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
 CC disease, myotonic dystrophy, sickle cell anaemia, thalassaemia,
 CC autoimmune/inflammatory disorders such as acquired immune deficiency
 CC syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
 CC atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,
 CC glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
 CC osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
 CC bacterial, fungal, parasitic, protozoal and helminthic infections, and
 CC cell proliferative disorders such as actinic keratosis, arteriosclerosis
 CC and cancer including breast, bladder, bone marrow, brain and uterus
 CC cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.
 XX
 XX Sequence 374 AA;

Query Match 100.0%; Score 1106; DB 22; Length 374;
 Best Local Similarity 100.0%; Pred. No. 2.5e-104;
 Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLSGQPVCRGGTQPCVKYIYFHDTSRRLNFEAEACRRDGGQQLVSI ESDQKLE 60
 Db 24 GRLLSGQPVCRGGTQPCVKYIYFHDTSRRLNFEAEACRRDGGQQLVSI ESDQKLE 83
 QY 61 KFIENLLPSDGDGFWIGLRREEKQSNSTACQDLYAWTDGSIQFRNWTYDEPSCGSEVCV 120
 Db 84 KFIENLLPSDGDGFWIGLRREEKQSNSTACQDLYAWTDGSIQFRNWTYDEPSCGSEVCV 143
 QY 121 VMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPVLP 180
 Db 144 VMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPVLP 203
 QY 181 BETQEDAKKTFKESREALNLAY 204
 Db 204 BETQEDAKKTFKESREALNLAY 227

RESULT 3

ABG66680
 ID ABG66680 standard; Protein; 374 AA.

AC ABG66680;

DT 30-AUG-2002 (first entry)

DE Human novel polypeptide #15.

Human; inflammatory condition; shock; sepsis; immune response;
 cancer; wound healing; central nervous system disease; haematopoiesis;
 peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
 myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
 cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
 bone degenerative disorder; periodontal disease; reperfusion injury;
 lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
 allergic condition; thrombolytic; thrombosis; coagulation disorder;
 fungal infection.

OS Homo sapiens.

PN WO200244340-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US47004.

XX 30-NOV-2000; 2000US-0028952.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
 PI Yamazaki V, Ujwal ML, Drmanac RT;

XX WPI; 2002-508509/54.

XX N-PSDB; ABK94904.

XX Novel nucleic acids and polypeptides for diagnosis, treatment of
 PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
 PT disorders, cancer and promoting wound healing -

PS Claim 10; Page 579-580; 672pp; English.

XX The invention relates to human novel polynucleotides and associated
 CC polypeptides. The polynucleotides and polypeptides are useful for
 CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
 CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
 CC and cancer and for promoting wound healing. The sequences are used to
 CC induce the proliferation of neural cells and regeneration of nerve and
 CC brain tissue, and are useful for the treatment of central and peripheral
 CC nervous system diseases and neuropathies, such as Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease and amyotrophic lateral
 CC sclerosis. The sequences are involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid

CC cell disorders and platelet disorders such as thrombocytopenia,
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
 CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
 CC disease. The sequences of the invention are also useful for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues, immune deficiencies and disorders
 CC including severe combined immunodeficiency (SCID), bacterial or fungal
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
 CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic
 CC and coagulation disorders. Sequences ABG66680-ABG66758 represent human
 CC novel polypeptides of the invention.

XX Sequence 374 AA;

Query Match 100.0%; Score 1106; DB 23; Length 374;
 Best Local Similarity 100.0%; Pred. No. 2.5e-104;
 Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLSGQPVCRGGTQPCVKYIYFHDTSRRLNFEAEACRRDGGQQLVSI ESDQKLE 60
 Db 24 GRLLSGQPVCRGGTQPCVKYIYFHDTSRRLNFEAEACRRDGGQQLVSI ESDQKLE 83
 QY 61 KFIENLLPSDGDGFWIGLRREEKQSNSTACQDLYAWTDGSIQFRNWTYDEPSCGSEVCV 120
 Db 84 KFIENLLPSDGDGFWIGLRREEKQSNSTACQDLYAWTDGSIQFRNWTYDEPSCGSEVCV 143
 QY 121 VMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPVLP 180
 Db 144 VMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSDEKPAVPSREAGEETELTPVLP 203
 QY 181 BETQEDAKKTFKESREALNLAY 204
 Db 204 BETQEDAKKTFKESREALNLAY 227

RESULT 4

ABG90203
 ID ABB90203 standard; Protein; 374 AA.

AC ABB90203;

XX 24-MAY-2002 (first entry)

XX Human polypeptide SEQ ID NO 2579.

XX Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.

OS Homo sapiens.

XX WO200190304-A2.

XX 29-NOV-2001.

XX 18-MAY-2001; 2001WO-US16450.

XX 19-MAY-2000; 2000US-205515P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX WPI; 2002-122018/16.

XX N-PSDB; ABL90612.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -

XX PS Claim 11; SEQ ID NO 2579; 2081pp + Sequence Listing; English.

XX XX The invention relates to novel genes (ABL9449-ABL90853) and proteins (ABB9040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

XX CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 374 AA;

Query Match 100.0%; Score 1106; DB 23; Length 374;
 Best Local Similarity 100.0%; Pred. No. 2.5e-104;
 Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLSQPVCRGCTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSEDEQKLE 60
 Db 24 GRLLSQPVCRGCTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSEDEQKLE 83

QY 61 KFTENLLPSDGDGFWIGLRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEVCV 120
 Db 84 KFTENLLPSDGDGFWIGLRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEVCV 143

QY 121 VMYHQSPAPAGIGGPPYMFQWDDRCNKNFNICKYSDEKPAVPSRAEGSETELTTPVLP 180
 Db 144 VMYHQSPAPAGIGGPPYMFQWDDRCNKNFNICKYSDEKPAVPSRAEGSETELTTPVLP 203

QY 181 EETQEDAKKTFKESREAAALNLAY 204
 Db 204 EETQEDAKKTFKESREAAALNLAY 227

RESULT 5
 AAM25796
 ID AAM25796 standard; Protein; 387 AA.
 XX AC AAM25796;
 XX XX 16-OCT-2001 (first entry)
 XX DE Human protein sequence SEQ ID NO:1311.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
 KW antiaggregant; haemostatic; vulnerary; antitumor; osteopathic; eczema;
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
 KW neurological disorder.

XX OS Homo sapiens.

XX XX

PN WO200153455-A2.

XX 26-JUL-2001.

XX 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457603/49.

XX N-PSDB; AAM99737.

XX Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX Claim 20; Page 272; 1217pp; English.

XX AAM99166 to AAM99904 encode the human proteins given in AAM25225 to AAM25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary; antitumor; osteopathic; dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmunity, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders.

XX SQ Sequence 387 AA;

Query Match 100.0%; Score 1106; DB 22; Length 387;
 Best Local Similarity 100.0%; Pred. No. 2.6e-104;
 Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRLLSQPVCRGCTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSEDEQKLE 60
 Db 37 GRLLSQPVCRGCTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSEDEQKLE 96

QY 61 KFTENLLPSDGDGFWIGLRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEVCV 120
 Db 97 KFTENLLPSDGDGFWIGLRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEVCV 156

QY 121 VMYHQSPAPAGIGGPPYMFQWDDRCNKNFNICKYSDEKPAVPSRAEGSETELTTPVLP 180
 Db 157 VMYHQSPAPAGIGGPPYMFQWDDRCNKNFNICKYSDEKPAVPSRAEGSETELTTPVLP 216

QY 181 EETQEDAKKTFKESREAAALNLAY 204
 Db 217 EETQEDAKKTFKESREAAALNLAY 240

RESULT 6
 AAY91490
 ID AAY91490 standard; Protein; 374 AA.
 XX AC AAY91490;
 XX XX

DT 29-JUN-2000 (first entry)
 XX Human secreted protein sequence encoded by gene 40 SEQ ID NO:163.
 DE Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 XX antiHIV; antiinflammatory; neutrotropic; neuroprotective; antiallergic;
 KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
 KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
 KW immune disease; inflammation; blood disorder; tumour.
 XX Homo sapiens.
 OS WO200006698-A1.
 PN 10-FEB-2000.
 XX 29-JUL-1999; 99WO-US17130.
 XX 30-JUL-1998; 98US-0094657.
 PR 05-AUG-1998; 98US-0095486.
 PR 06-AUG-1998; 98US-0095454.
 PR 06-AUG-1998; 98US-0095455.
 PR 12-AUG-1998; 98US-0096319.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Komatsoulis GA, Rosen SM, Duan R, Moore PA, Shi Y;
 PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
 PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
 DR WPI; 2000-195282/17.
 DR N-PSDB; AAA26385.
 XX New isolated human genes and the secreted polypeptides they encode,
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders -
 XX Claim 11; Page 483-484; 634pp; English.
 XX The polynucleotide sequences given in AAA26346 to AAA26458 encode the
 CC human secreted proteins given in AAY91451 to AAY91691. The human secreted
 CC proteins can have activities based on the tissues and cells they are
 CC expressed in. Examples of the activities are: cytostatic;
 CC immunosuppressive; antiHIV; antiinflammatory; neutrotropic; neuroprotective;
 CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
 CC antiasthma; antipsoriatic; and cardiant. The polynucleotides and their
 CC corresponding secreted proteins are useful for preventing, treating or
 CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
 CC pathological conditions can be diagnosed by determining the amount of the
 CC proteins in a sample or by determining the presence of mutations in the
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, inflammation,
 CC allergies, Alzheimer's and behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, sepsis, acne, psoriasis,
 CC transplant rejection, diabetes, asthma, spinal cord injuries,
 CC cardiovascular disorders, reproductive disorders, gastrointestinal
 CC disorders, respiratory disorders and metabolic disorders. The proteins
 CC or polynucleotides can also be used as food additives or preservatives.
 CC The proteins are also useful for identifying their binding partners.
 CC AAA26337 to AAA26345 and AAY91450 are sequences used in the
 CC exemplification of the present invention.
 XX Sequence 374 AA;
 SQ
 Query Match 99.1%; Score 1096; DB 21; Length 374;
 Best Local Similarity 99.0%; Pred. No. 2.6e-103;
 Matches 202; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GRLLSGQPVCRGQTQRCYKVIYFHDTSRLNPFEEAKEACRRDGGQLVSI ESEDEQKLE 60

Db 24 GRLLSGQPVCRGQTQRCYKVIYFHDTSRLNPFEEAKEACRRDGGQLVSI ESEDEQKLE 83
 Qy 61 KPTIENLLPSGDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNWTVDPSGSEVCV 120
 Db 84 XFTIENLLPSGDGFWIGLRRREKQSNSTXKQDLYAWTDGSIQFRNWTVDPSGSEVCV 143
 Qy 121 VMYHQPAPAGIGGPPYMFQNDRCNNKNNFICKYSDPKPAVPSREAGBETELTTPVLP 180
 Db 144 VMYHQPAPAGIGGPPYMFQNDRCNNKNNFICKYSDPKPAVPSREAGBETELTTPVLP 203
 Qy 181 EETQEDAKKTFKESREAAALNAY 204
 Db 204 EETQEDAKKTFKESREAAALNAY 227
 RESULT 7
 AAY13367
 ID AAY13367 standard; Protein; 382 AA.
 XX
 AC AAY13367;
 XX
 DT 25-JUN-1999 (first entry)
 XX
 DE Amino acid sequence of protein PRO234.
 XX
 KW Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW congenital microvillus atrophy; skin disease; cell growth;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
 KW anti-thrombotic; wound healing; tissue repair.
 XX
 OS Homo sapiens.
 XX
 PN WO9914328-A2.
 XX
 PD 25-MAR-1999.
 XX
 PF 16-SEP-1998; 98WO-US19330.
 XX
 PR 25-NOV-1997; 97US-0066840.
 PR 17-SEP-1997; 97US-0059113.
 PR 17-SEP-1997; 97US-0059115.
 PR 17-SEP-1997; 97US-0059117.
 PR 17-SEP-1997; 97US-0059119.
 PR 17-SEP-1997; 97US-0059121.
 PR 17-SEP-1997; 97US-0059122.
 PR 17-SEP-1997; 97US-0059184.
 PR 18-SEP-1997; 97US-0059263.
 PR 18-SEP-1997; 97US-0059266.
 PR 15-OCT-1997; 97US-0062125.
 PR 17-OCT-1997; 97US-0062285.
 PR 17-OCT-1997; 97US-0062287.
 PR 21-OCT-1997; 97US-0063486.
 PR 24-OCT-1997; 97US-0062814.
 PR 24-OCT-1997; 97US-0062816.
 PR 24-OCT-1997; 97US-0063045.
 PR 24-OCT-1997; 97US-0063120.
 PR 24-OCT-1997; 97US-0063121.
 PR 24-OCT-1997; 97US-0063127.
 PR 24-OCT-1997; 97US-0063128.
 PR 27-OCT-1997; 97US-0063329.
 PR 27-OCT-1997; 97US-0063327.
 PR 28-OCT-1997; 97US-0063541.
 PR 28-OCT-1997; 97US-0063542.
 PR 28-OCT-1997; 97US-0063544.
 PR 28-OCT-1997; 97US-0063549.
 PR 28-OCT-1997; 97US-0063550.
 PR 29-OCT-1997; 97US-0063564.
 PR 29-OCT-1997; 97US-0063435.
 PR 29-OCT-1997; 97US-0063704.

PR 29-OCT-1997; 97US-0063732.
 PR 29-OCT-1997; 97US-0063738.
 PR 29-OCT-1997; 97US-0063734.
 PR 29-OCT-1997; 97US-0064215.
 PR 29-OCT-1997; 97US-0063735.
 PR 31-OCT-1997; 97US-0063870.
 PR 31-OCT-1997; 97US-0064103.
 PR 03-NOV-1997; 97US-0064248.
 PR 07-NOV-1997; 97US-0064809.
 PR 12-NOV-1997; 97US-0065186.
 PR 17-NOV-1997; 97US-0065846.
 PR 18-NOV-1997; 97US-0065693.
 PR 21-NOV-1997; 97US-0066120.
 PR 21-NOV-1997; 97US-0066364.
 PR 24-NOV-1997; 97US-0066772.
 PR 24-NOV-1997; 97US-0066466.
 PR 24-NOV-1997; 97US-0066770.
 PR 24-NOV-1997; 97US-0066511.
 PR 24-NOV-1997; 97US-0066453.
 XX (GETH) GENENTECH INC.
 PA
 XX
 XX
 XX
 PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 XX WPI; 1999-229533/19.
 DR N-PSDB; AAX52238.
 XX
 XX
 PT New isolated human genes and polypeptides used in, e.g. treatment of
 PT gastrointestinal ulceration
 XX
 XX
 XX Claim 12; Fig 50; 320pp; English.
 XX
 CC AAY13344-403 represent secreted and transmembrane human proteins.
 CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to
 CC growth or survival of nerve cells including Parkinson's disease,
 CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
 CC for fibromodulin, e.g. for reducing dermal scarring. PRO265 can be used as
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair; PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.
 XX
 XX
 SQ Sequence 382 AA;
 Query Match 98.7%; Score 1092; DB 20; Length 382;
 Best Local Similarity 96.2%; Pred. No. 7e-103;
 Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 GRLLS-----GQPVCRGRTQPCVKYIFHDTSRRLNFEAKACRRDGGQLVSIES 52
 DB 24 GRLLSASDLRLGQPVCRGRTQPCVKYIFHDTSRRLNFEAKACRRDGGQLVSIES 83
 QY 53 EDEQKLIETFIENLLPSDGDFTWGLRRREKQSNSTACQDLYAWTQSGISQFRNWTYVDEP 112
 DB 84 EDEQKLIETFIENLLPSDGDFTWGLRRREKQSNSTACQDLYAWTQSGISQFRNWTYVDEP 143
 QY 113 SCGSEVCVVMYHOPSPAGIGGYPFMQWNDRCNMKNFICKYSDSKPAVPSREARGEET 172
 DB 144 SCGSEVCVVMYHOPSPAGIGGYPFMQWNDRCNMKNFICKYSDSKPAVPSREARGEET 203
 QY 173 ELTTPVLPBETQEDAKTKFKESREAAALNLAY 204

Db 204 ELTTPVLPBETQEDAKTKFKESREAAALNLAY 235
 RESULT 8
 AAU29033
 ID AAU29033 standard; Protein; 382 AA.
 XX
 AC AAU29033;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human PRO polypeptide sequence #10.
 XX
 KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200168848-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06520.
 XX
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 03-MAR-2000; 2000US-187202P.
 PR 06-MAR-2000; 2000US-186968P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 14-MAR-2000; 2000US-189328P.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 21-MAR-2000; 2000US-190828P.
 PR 21-MAR-2000; 2000US-191007P.
 PR 21-MAR-2000; 2000US-191048P.
 PR 21-MAR-2000; 2000US-191314P.
 PR 28-MAR-2000; 2000US-192655P.
 PR 29-MAR-2000; 2000US-193032P.
 PR 30-MAR-2000; 2000US-193053P.
 PR 04-APR-2000; 2000WO-US08439.
 PR 04-APR-2000; 2000US-194449P.
 PR 11-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 18-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 25-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 03-MAY-2000; 2000US-199654P.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 22-AUG-2000; 2000US-0644848.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 (GETH) GENENTECH INC.
 PA
 XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Fan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX WPI; 2001-602746/68.
 DR


```

DR N-PSDB; AAS45934.
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX
XX Claim 11; Fig 20; 77app; English.
XX
XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
XX The PRO polypeptides and their associated nucleic acids can be used to
XX detect the presence of a tumour in a mammal by comparing the level of
XX expression of a PRO polypeptide in a test sample of cells from the animal
XX and a control sample of normal cells, whereby a higher level of
XX expression in the test sample indicates the presence of a tumour in the
XX mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
XX and rabbits but are preferably human. The polypeptides can be used to
XX stimulate tumour necrosis factor (TNF) alpha release from human blood,
XX when contacted with it. A specific polypeptide can be used to stimulate
XX the proliferation or differentiation of chondrocyte cells. The PRO
XX proteins can be used to determine the presence of tumours and also
XX susceptibility to tumour development, particularly adrenal, lung, colon,
XX breast, prostate, rectal, cervical, or liver tumours, in mammalian
XX subjects. The oligonucleotide probes specific for the PRO nucleic acids
XX can be used for genetic analysis of individuals with genetic disorders.
XX
XX Sequence 382 AA;
XX
XX Query Match          98.7%; Score 1092; DB 22; Length 382;
XX Best Local Similarity 96.2%; Pred. No. 7e-103;
XX Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
XX
QY 1 GRLLS-----GQPVCRGGTQPCYKVIYFHDTSRLNFEFEAKEACRRDGGQLVSIES 52
DB 24 GRLLSASDLRLGGGQPVCRGGTQPCYKVIYFHDTSRLNFEFEAKEACRRDGGQLVSIES 83
QY 53 EDEOKLIEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEP 112
DB 84 EDEOKLIEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEP 143
QY 113 SCGSEVCVMYHOPSAPAGIGGPFYMFQWDDRCNMKNFNICKYSDEKPAVPSREAGBET 172
DB 144 SCGSEVCVMYHOPSAPAGIGGPFYMFQWDDRCNMKNFNICKYSDEKPAVPSREAGBET 203
QY 173 ELTTPVLPEETQEDAKKTFKESREAAALNLAY 204
DB 204 ELTTPVLPEETQEDAKKTFKESREAAALNLAY 235
XX
XX RESULT 9
XX AAB80235
XX ID AAB80235 standard; Protein; 382 AA.
XX
XX AC AAB80235;
XX
XX DT 24-APR-2001 (first entry)
XX
XX DE Human PRO234 protein.
XX
XX KW Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
XX antiparkinsonian nootropic; neuroprotective; vulnary; cardiac;
XX angiogenic; vasotropic; antisthmatic; antirheumatic; cancer;
XX antiarthritic; antiinfertility; antidiabetic; antiviral; diabetes;
XX ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
XX ischaemia; inflammation.
XX
XX OS Homo sapiens.
XX
XX PN WO200104311-A1.
XX
XX PD 18-JAN-2001.
XX
XX PF 22-FEB-2000; 2000WO-US04414.
XX
XX

```

```

PR 07-JUL-1999; 99US-0143048.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20594.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 29-NOV-1999; 99WO-US28214.
PR 30-NOV-1999; 99WO-US28313.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 05-JAN-2000; 99WO-US00219.
XX
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
XX Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
XX Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kijavini IJ;
XX Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
XX Williams FM, Wood WI;
XX
XX WPI; 2001-081051/09.
XX N-PSDB; AAF72396.
XX
XX Sixty one nucleic acids encoding PRO polypeptides which are useful in
XX the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
XX squamous cell carcinoma) and neurodegenerative diseases (e.g.
XX Alzheimer's disease) -
XX
XX Claim 1; Fig 50; 393pp; English.
XX
XX The present sequence is one of sixty one novel secreted and
XX transmembrane PRO polypeptides. The PRO polypeptides are
XX useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
XX squamous cell carcinoma), gastrointestinal disorders (e.g.
XX enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
XX Parkinson's disease), wound repair, cardiovascular disorders (e.g.
XX endometrial bleeding angiogenesis, ischaemias such as coronary
XX ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
XX rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
XX diabetes and retinal disorders such as retinitis pigmentosa.
XX The PRO nucleic acids have applications in molecular biology, including
XX use as hybridization probes, and in chromosome and gene mapping.
XX
XX Sequence 382 AA;
XX
XX Query Match          98.7%; Score 1092; DB 22; Length 382;
XX Best Local Similarity 96.2%; Pred. No. 7e-103;
XX Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
XX
QY 1 GRLLS-----GQPVCRGGTQPCYKVIYFHDTSRLNFEFEAKEACRRDGGQLVSIES 52
DB 24 GRLLSASDLRLGGGQPVCRGGTQPCYKVIYFHDTSRLNFEFEAKEACRRDGGQLVSIES 83
QY 53 EDEOKLIEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEP 112
DB 84 EDEOKLIEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEP 143
QY 113 SCGSEVCVMYHOPSAPAGIGGPFYMFQWDDRCNMKNFNICKYSDEKPAVPSREAGBET 172
DB 144 SCGSEVCVMYHOPSAPAGIGGPFYMFQWDDRCNMKNFNICKYSDEKPAVPSREAGBET 203
QY 173 ELTTPVLPEETQEDAKKTFKESREAAALNLAY 204
DB 204 ELTTPVLPEETQEDAKKTFKESREAAALNLAY 235
XX
XX RESULT 10
XX ABU69645
XX ID ABU69645 standard; Protein; 382 AA.
XX

```


CC protein electrophoresis purposes. The anti-PRO antibodies may be used
CC in diagnostic assays for PRO, or for the affinity purification of PRO
CC from recombinant cell culture or natural sources. This is the amino
CC acid sequence of a novel human PRO polypeptide.

XX Sequence 382 AA;

SQ Query Match 98.7%; Score 1092; DB 24; Length 382;
Best Local Similarity 96.2%; Pred. No. 7e-103;
Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 GRLLS-----CQPCRGCTQPCYKVIYFHDTSRLNFEFAKEACRRDGGQLVSI 52
Db 24 GRLLSADLDLRGQPCRGCTQPCYKVIYFHDTSRLNFEFAKEACRRDGGQLVSI 83
QY 53 EDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEP 112
Db 84 EDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEP 143
QY 113 SCGSEVCVMYHQSAPAGIGGPTMFQWDDRCNMKNFICKYSDEKPAVPSREAEGEET 172
Db 144 SCGSEVCVMYHQSAPAGIGGPTMFQWDDRCNMKNFICKYSDEKPAVPSREAEGEET 203
QY 173 ELTTPVLPETQEDAKKTKESREAALNLAY 204
Db 204 ELTTPVLPETQEDAKKTKESREAALNLAY 235

RESULT 11

ABU71121
ID ABU71121 standard; Protein; 382 AA.

XX AC ABU71121;

XX DT 10-JUN-2003 (first entry)

XX DE Human PRO234 protein.

XX KW Human; PRO; secreted; transmembrane; cytostatic; TNF-alpha; blood;
KW tumour necrosis factor alpha release; chondrocyte cell; proliferation;
KW differentiation; tumour; gene therapy.

XX OS Homo sapiens.

XX PN UC2003036143-A1.

XX PD 20-FEB-2003.

XX PF 02-JUL-2002; 2002US-0187600.

XX PR 16-SEP-1998; 98WO-US19330.

PR 07-OCT-1998; 98WO-US21141.

PR 01-DEC-1998; 98WO-US25108.

PR 08-MAR-1999; 99WO-US05028.

PR 14-MAY-1999; 99WO-US10733.

PR 02-JUN-1999; 99WO-US12252.

PR 01-SEP-1999; 99WO-US20111.

PR 01-SEP-1999; 99WO-US21090.

PR 02-DEC-1999; 99WO-US28301.

PR 30-DEC-1999; 99WO-US31274.

PR 05-JAN-2000; 2000WO-US00219.

PR 18-FEB-2000; 2000WO-US04341.

PR 22-FEB-2000; 2000WO-US04342.

PR 24-FEB-2000; 2000WO-US04414.

PR 01-MAR-2000; 2000WO-US05601.

PR 02-MAR-2000; 2000WO-US05601.

PR 15-MAR-2000; 2000WO-US05684.

PR 30-MAR-2000; 2000WO-US08439.

PR 17-MAY-2000; 2000WO-US13705.

PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000WO-US14941.

PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 29-AUG-2001; 2001WO-US27099.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 29-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069425P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083495P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 15-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085560P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.

```
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088038P.
PR 04-JUN-1998; 98US-088039P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088036P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 05-JUN-1998; 98US-088655P.
PR 05-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096871P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 26-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.

PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.
PR 01-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.
PR 10-SEP-1998; 98US-099754P.
PR 10-SEP-1998; 98US-099763P.
PR 10-SEP-1998; 98US-099812P.

Query Match 98.7%; Score 1092; DB 24; Length 382;
Best Local Similarity 96.2%; Pred. No. 7e-103;
Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 GRLLS-----GQVCRGGTQPCYKVIYFHTSRRLNFEAKACRRDGGQLVSI 52
Db 24 GRLLSASDLRLGGQPCVCRGGTQPCYKVIYFHTSRRLNFEAKACRRDGGQLVSI 83
QY 53 EDEOKLIEKFIENLLPSDGFWMGLRRRREKOSNSTACODLYAWTSGISQFRNMYVDEP 112
Db 84 EDEOKLIEKFIENLLPSDGFWMGLRRRREKOSNSTACODLYAWTSGISQFRNMYVDEP 143
QY 113 SCGSEVCVVMYHQPAPAGIGGYPMPQWMDRCNMKNFNICKYSDEKPAVPSREASGEET 172
Db 144 SCGSEVCVVMYHQPAPAGIGGYPMPQWMDRCNMKNFNICKYSDEKPAVPSREASGEET 203
QY 173 ELTTPVLPEETQEDAKTKFKSREAAALNLAY 204
Db 204 ELTTPVLPEETQEDAKTKFKSREAAALNLAY 235

RESULT 12
ABU71468
ID ABU71468 standard; Protein; 382 AA.
XX
AC ABU71468;
XX
DT 10-JUN-2003 (first entry)
XX
DE Human PRO polypeptide #24.
XX
KW Human; secreted and transmembrane protein; PRO polypeptide; cancer;
KW Alzheimer's disease; ischaemia; cytostatic; neurotropic; vasotropic;
KW neuroprotective.
XX
OS Homo sapiens.
XX
PN US2002192659-A1.
XX
PD 19-DEC-2002.
XX
PF 10-JUL-2001; 2001US-0902853.
XX
PR 10-SEP-1998; 98WO-US18824.
PR 14-SEP-1998; 98WO-US19177.
PR 16-SEP-1998; 98WO-US19330.
PR 17-SEP-1998; 98WO-US19437.
PR 01-DEC-1998; 98WO-US25108.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28564.
PR 16-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 05-JAN-2000; 2000WO-US00219.
PR 11-FEB-2000; 2000WO-US03565.
```

```

PR 22-FEB-2000; 200WO-US04414.
PR 28-JUL-2000; 200WO-US20710.
PR 24-AUG-2000; 200WO-US23328.
PR 17-SEP-1997; 97US-059113P.
PR 17-SEP-1997; 97US-059115P.
PR 18-SEP-1997; 97US-059266P.
PR 15-OCT-1997; 97US-062125P.
PR 17-OCT-1997; 97US-062285P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-062814P.
PR 24-OCT-1997; 97US-062816P.
XX PA (GETH ) GENENTECH INC.
XX PI Ashkenazi A, Botstein D, Desnovers L, Eaton DL, Ferrara N;
PI Pilvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kljavin IJ;
PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
PI Williams PM, Wood WI;
XX DR WPI; 2003-361832/34.
XX DR N-PSDB; ACAS8405.
XX PT New isolated nucleic acid encoding a PRO polypeptide, e.g. PRO245 or
XX PT PRO1868, useful in molecular biology, chromosome and gene mapping, in
XX PT generating antisense RNA and DNA, and in gene therapy -
XX PS Claim 12; Fig 50; 474pp; English.
XX CC The present invention relates to the isolation of novel human secreted
XX CC and transmembrane proteins (PRO polypeptides), and the polynucleotide
XX CC sequences encoding them. The polynucleotide sequences are useful in
XX CC molecular biology, as hybridization probes, in chromosome and gene
XX CC mapping, in generating antisense RNA and DNA, and in gene therapy. The
XX CC polynucleotide sequences may also be used in preparing PRO polypeptides
XX CC by recombinant techniques, and in generating either transgenic animals
XX CC or knock-out animals which, in turn, are useful in the development and
XX CC screening of therapeutically useful reagents. The PRO polypeptides or
XX CC their antibodies are useful in preparing a medicament for treating a
XX CC condition responsive to the polypeptide or antibody, such as cancer,
XX CC Alzheimer's disease or ischaemia, and in various diagnostic assays.
XX CC ABU71445-ABU71505 represent human PRO polypeptides of the invention.
XX SQ Sequence 382 AA;

Query Match 98.7%; Score 1092; DB 24; Length 382;
Best Local Similarity 96.2%; Pred. No. 7e-103;
Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 GRLLS-----GQPVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSI 52
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 53 EDSQKLEIEIENLLPSDGFHWGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEP 112
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 84 EDSQKLEIEIENLLPSDGFHWGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEP 143
QY 113 SCGSEVCVMYHOPSPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAGEET 172
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 144 SCGSEVCVMYHOPSPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAGEET 203
QY 173 ELTPVLPETQEDAKTKFKESREALNLAY 204
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 204 ELTPVLPETQEDAKTKFKESREALNLAY 235

RESULT 13
ABU71914
ID ABU71914 standard; Protein; 382 AA.
XX
AC ABU71914;

```

```

XX DT 12-JUN-2003 (first entry)
XX DE Human secreted/transmembrane protein PRO234.
XX KW Human; secreted protein; transmembrane protein; PRO;
XX KW gene therapy; chromosome identification; chromosome marker.
XX OS Homo sapiens.
XX PN US2003003530-A1.
XX PD 02-JAN-2003.
XX PF 11-JUL-2001; 2001US-0904011.
XX PR 10-SEP-1998; 98WO-US18824.
XX PR 14-SEP-1998; 98WO-US19177.
XX PR 16-SEP-1998; 98WO-US19330.
XX PR 17-SEP-1998; 98WO-US19437.
XX PR 01-DEC-1998; 98WO-US25108.
XX PR 08-SEP-1999; 99WO-US20594.
XX PR 13-SEP-1999; 99WO-US20944.
XX PR 15-SEP-1999; 99WO-US21090.
XX PR 15-SEP-1999; 99WO-US21547.
XX PR 05-OCT-1999; 99WO-US21089.
XX PR 29-NOV-1999; 99WO-US28214.
XX PR 30-NOV-1999; 99WO-US28313.
XX PR 01-DEC-1999; 99WO-US28301.
XX PR 02-DEC-1999; 99WO-US28564.
XX PR 16-DEC-1999; 99WO-US28565.
XX PR 20-DEC-1999; 99WO-US30095.
XX PR 20-DEC-1999; 99WO-US30911.
XX PR 20-DEC-1999; 99WO-US30999.
XX PR 05-JAN-2000; 2000WO-US00219.
XX PR 11-FEB-2000; 2000WO-US03565.
XX PR 22-FEB-2000; 2000WO-US04414.
XX PR 24-FEB-2000; 2000WO-US05004.
XX PR 02-MAR-2000; 2000WO-US05841.
XX PR 20-MAR-2000; 2000WO-US07377.
XX PR 30-MAR-2000; 2000WO-US08439.
XX PR 02-MAY-2000; 2000WO-US14042.
XX PR 02-JUN-2000; 2000WO-US15264.
XX PR 28-JUL-2000; 2000WO-US20710.
XX PR 24-AUG-2000; 2000WO-US23328.
XX PR 17-SEP-1997; 97US-059113P.
XX PR 17-SEP-1997; 97US-059115P.
XX PR 17-SEP-1997; 97US-059117P.
XX PR 17-SEP-1997; 97US-059119P.
XX PR 17-SEP-1997; 97US-059121P.
XX PR 17-SEP-1997; 97US-059122P.
XX PR 18-SEP-1997; 97US-059184P.
XX PR 18-SEP-1997; 97US-059263P.
XX PR 18-SEP-1997; 97US-059266P.
XX PR 15-OCT-1997; 97US-062125P.
XX PR 17-OCT-1997; 97US-062285P.
XX PR 17-OCT-1997; 97US-062287P.
XX PR 21-OCT-1997; 97US-063486P.
XX PR 24-OCT-1997; 97US-062814P.
XX PR 24-OCT-1997; 97US-062816P.
XX PR 24-OCT-1997; 97US-063045P.
XX PR 24-OCT-1997; 97US-063120P.
XX PR 24-OCT-1997; 97US-063121P.
XX PR 24-OCT-1997; 97US-063127P.
XX PR 24-OCT-1997; 97US-063128P.
XX PR 27-OCT-1997; 97US-063327P.
XX PR 27-OCT-1997; 97US-063329P.
XX PR 28-OCT-1997; 97US-063541P.
XX PR 28-OCT-1997; 97US-063542P.
XX PR 28-OCT-1997; 97US-063544P.
XX PR 28-OCT-1997; 97US-063549P.
XX PR 28-OCT-1997; 97US-063550P.
XX PR 28-OCT-1997; 97US-063564P.

```

PR 29-OCT-1997; 97US-0634435P.
 PR 29-OCT-1997; 97US-063704P.
 PR 29-OCT-1997; 97US-063732P.
 PR 29-OCT-1997; 97US-063734P.
 PR 29-OCT-1997; 97US-063735P.
 PR 29-OCT-1997; 97US-063738P.
 PR 29-OCT-1997; 97US-064215P.
 PR 31-OCT-1997; 97US-063870P.
 PR 31-OCT-1997; 97US-064103P.
 PR 03-NOV-1997; 97US-064248P.
 PR 07-NOV-1997; 97US-064809P.
 PR 12-NOV-1997; 97US-065186P.
 PR 17-NOV-1997; 97US-065846P.
 PR 18-NOV-1997; 97US-065693P.
 PR 21-NOV-1997; 97US-066120P.
 PR 24-NOV-1997; 97US-066364P.
 PR 24-NOV-1997; 97US-066453P.
 PR 24-NOV-1997; 97US-066466P.
 PR 24-NOV-1997; 97US-066511P.
 PR 24-NOV-1997; 97US-066770P.
 PR 18-SEP-2000; 2000US-0665350.
 XX
 PA (GETH) GENENTECH INC.

XX Ashkenazi A, Botstein D, Desnovers L, Eaton DL, Ferrara N;
 PI Filvaroff E, Pong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kljavin LJ;
 PI Mather JP, Pan J, Pooni NF, Roy MA, Stewart TA, Tumas D;
 PI Williams PM, Wood WI;
 XX
 DR WPI; 2003-329602/31.
 DR N-PSDB; ACA60112.

XX New transmembrane polypeptides and nucleic acids encoding the
 PT polypeptides, useful in gene therapy, in chromosome identification, as
 PT chromosome markers, in generating probes and in tissue typing -

XX Claim 12; Fig 50; 484pp; English.

XX The invention relates to an isolated nucleic acid with at least 80%
 CC nucleic acid sequence identity to a nucleotide sequence encoding one of
 CC 61 secreted/transmembrane polypeptides, or PRO polypeptides or encoding a
 CC PRO protein extracellular domain. Also included are a vector comprising
 CC the PRO nucleic acid, a host cell comprising the vector, producing a PRO
 CC polypeptide (by culturing the host cell for the expression of the PRO
 CC polypeptide, and recovering the PRO polypeptide from the cell culture),
 CC an isolated PRO polypeptide (having at least 80% sequence identity
 CC to: (a) an amino acid sequence selected from the 61 PRO proteins;
 CC (b) an amino acid sequence encoded by a nucleic acid molecule deposited
 CC with an ATCC number (detailed in the specification); or (c) an
 CC extracellular domain of a PRO polypeptide or to a PRO polypeptide lacking
 CC its associated signal peptide), a chimeric molecule comprising a PRO
 CC polypeptide of fused to a heterologous amino acid sequence, an anti-PRO
 CC antibody, detecting a PRO245 or PRO1868 in a sample suspected of
 CC containing the polypeptide, linking a bioactive molecule to a cell
 CC expressing a PRO245 or PRO1868 and modulating at least one biological
 CC activity of a cell expressing a PRO245 or PRO1868. Nucleic acids which
 CC encode PRO can be used to generate either transgenic animals or knock-out
 CC animals which may be used in the development and screening of
 CC therapeutically useful reagents. The nucleic acids may also be used in
 CC gene therapy, in chromosome identification, as chromosome markers, or in
 CC generating probes. The PRO polypeptides are useful as molecular markers
 CC for protein electrophoresis, and the isolated nucleic acids may be used
 CC for recombinantly expressing those markers. The PRO polypeptides and
 CC nucleic acids may also be used in tissue typing. Anti-PRO antibodies
 CC are useful in diagnostic assays for PRO, and in affinity purification
 CC of PRO from recombinant cell culture or natural sources. The
 CC present sequence represents a PRO protein.

XX Sequence 382 AA;

Query Match 98.7%; Score 1092; DB 24; Length 382;

Best Local Similarity 96.2%; Pred. No. 7e-103;
 Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 GRLLS-----GQPVCRGGTQRCYKVIYFHDTSRRLNFEAKACRRDGGQLVSIES 52
 DB 24 GRLLSASDLRLGGQPVCRGGTQRCYKVIYFHDTSRRLNFEAKACRRDGGQLVSIES 83
 QY 53 EDEQKLIKFIENLLPSDGFWIGLRREEKQSNSTACODLYAWTDGSIQFERNWYVDEP 112
 DB 84 EDEQKLIKFIENLLPSDGFWIGLRREEKQSNSTACODLYAWTDGSIQFERNWYVDEP 143
 QY 113 SCGSEVCVVVYHQPSPAGIGGYPYMFQWMDRCNMKNFNICKYSDEKPAVPSREASGEET 172
 DB 144 SCGSEVCVVVYHQPSPAGIGGYPYMFQWMDRCNMKNFNICKYSDEKPAVPSREASGEET 203
 QY 173 ELTTPVLPEETQBEDAKTKFKESREAAALNLAY 204
 DB 204 ELTTPVLPEETQBEDAKTKFKESREAAALNLAY 235

RESULT 14

ABU65578
 ID ABU65578 standard; Protein; 382 AA.

XX AC ABU65578;

XX DT 19-MAY-2003 (first entry)

XX XX Human secreted/transmembrane protein, SEQ ID 20.

KW Human; PRO; secreted protein; transmembrane protein;
 KW cytostatic; antiarthritic; osteopathic; adrenal tumour; lung tumour;
 KW colon tumour; breast tumour; prostate tumour; rectal tumour;
 KW cervical tumour; liver tumour; TNF-alpha release; arthritis;
 KW tumour necrosis factor alpha; chondrocyte cell; bone disorder;
 KW cartilage disorder; sports injury.

XX OS Homo sapiens.

XX PN US2003036156-A1.

XX PD 20-FEB-2003.

XX PF 02-JUL-2002; 2002US-0188767.

XX PR 16-SEP-1998; 98WO-US19330.

PR 07-OCT-1998; 98WO-US21141.

PR 01-DEC-1998; 98WO-US25108.

PR 08-MAR-1999; 99WO-US05028.

PR 14-MAY-1999; 99WO-US10733.

PR 02-JUN-1999; 99WO-US12252.

PR 01-SEP-1999; 99WO-US20111.

PR 15-SEP-1999; 99WO-US21090.

PR 01-DEC-1999; 99WO-US28301.

PR 30-DEC-1999; 99WO-US31274.

PR 05-JAN-2000; 2000WO-US00219.

PR 18-FEB-2000; 2000WO-US04341.

PR 18-FEB-2000; 2000WO-US04342.

PR 22-FEB-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.

PR 02-MAR-2000; 2000WO-US05841.

PR 15-MAR-2000; 2000WO-US06884.

PR 30-MAR-2000; 2000WO-US08439.

PR 17-MAY-2000; 2000WO-US13705.

PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000WO-US14941.

PR 02-JUN-2000; 2000WO-US15264.

PR 28-JUL-2000; 2000WO-US20710.

PR 08-NOV-2000; 2000WO-US23328.

PR 01-DEC-2000; 2000WO-US32678.

PR 20-DEC-2000; 200OWO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 29-AUG-2001; 2001WO-US217099.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 29-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069425P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084633P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.

PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 11-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094005P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.
PR 01-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.

```
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.

Query Match 98.7%; Score 1092; DB 24; Length 382;
Best Local Similarity 96.2%; Pred. No. 7e-103;
Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 GRLLS-----GQPCVCGGTORPCYKVIYFHDTSRLNFEAKACRDGGQLYSIES 52
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24 GRLLSADLLRGQPCVCGGTORPCYKVIYFHDTSRLNFEAKACRDGGQLYSIES 83
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 53 EDSQKLIKFIENLLPSDGFWGLRRREKQSNSTACQDLYAWTDGSIQSFFNWTYVDEP 112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 84 EDSQKLIKFIENLLPSDGFWGLRRREKQSNSTACQDLYAWTDGSIQSFFNWTYVDEP 143
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 113 SCSEVCVVMYHOPSPAGIGGYPFMQWDDRCNMKNFICKYSDEKPAVPSREAGEET 172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 144 SCSEVCVVMYHOPSPAGIGGYPFMQWDDRCNMKNFICKYSDEKPAVPSREAGEET 203
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 173 ELTPVLPETQEDAKTKFKESREAALNLAY 204
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 204 ELTPVLPETQEDAKTKFKESREAALNLAY 235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
ABU65911
ID ABU65911 standard; Protein; 382 AA.
AC ABU65911;
XX
XX
XX DT 20-MAY-2003 (first entry)
XX
XX DE Novel human secreted and transmembrane protein PRO234.
XX
XX KW Human; secreted protein; transmembrane protein; cytostatic;
KW gene therapy; TNF-Agonist-Alpha; chondrocyte stimulator; tumour;
KW adrenal tumour; lung tumour; colon tumour; breast tumour;
KW prostate tumour; rectal tumour; cervical tumour; liver tumour.
XX
XX OS Homo sapiens.
XX
XX PN US2003036157-A1.
XX
XX PD 20-FEB-2003.
XX
XX PF 02-JUL-2002; 2002US-0188769.
XX
XX PR 16-SEP-1998; 98WO-US19330.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 08-MAR-1999; 99WO-US05028.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28551.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.

24-AUG-2000; 2000WO-US23328.
08-NOV-2000; 2000WO-US30952.
01-DEC-2000; 2000WO-US32678.
20-DEC-2000; 2000WO-US34956.
28-FEB-2001; 2001WO-US06520.
01-JUN-2001; 2001WO-US17800.
20-JUN-2001; 2001WO-US19692.
29-JUN-2001; 2001WO-US21066.
09-JUL-2001; 2001WO-US21735.
29-AUG-2001; 2001WO-US27099.
18-SEP-1997; 97US-059263P.
18-SEP-1997; 97US-059266P.
17-OCT-1997; 97US-062250P.
21-OCT-1997; 97US-063486P.
24-OCT-1997; 97US-063120P.
24-OCT-1997; 97US-063121P.
28-OCT-1997; 97US-063540P.
28-OCT-1997; 97US-063541P.
28-OCT-1997; 97US-063544P.
28-OCT-1997; 97US-063564P.
29-OCT-1997; 97US-063734P.
31-OCT-1997; 97US-063870P.
31-OCT-1997; 97US-064103P.
13-NOV-1997; 97US-065311P.
21-NOV-1997; 97US-066120P.
24-NOV-1997; 97US-066466P.
11-DEC-1997; 97US-066772P.
12-DEC-1997; 97US-069335P.
17-DEC-1997; 97US-069425P.
18-DEC-1997; 97US-069870P.
10-MAR-1998; 98US-068017P.
11-MAR-1998; 98US-077450P.
11-MAR-1998; 98US-077632P.
11-MAR-1998; 98US-077649P.
20-MAR-1998; 98US-078886P.
20-MAR-1998; 98US-078939P.
27-MAR-1998; 98US-079664P.
27-MAR-1998; 98US-079786P.
31-MAR-1998; 98US-080107P.
31-MAR-1998; 98US-080194P.
01-APR-1998; 98US-080327P.
01-APR-1998; 98US-080333P.
08-APR-1998; 98US-081049P.
08-APR-1998; 98US-081070P.
09-APR-1998; 98US-081195P.
15-APR-1998; 98US-081838P.
21-APR-1998; 98US-082568P.
21-APR-1998; 98US-082569P.
22-APR-1998; 98US-082704P.
22-APR-1998; 98US-082797P.
28-APR-1998; 98US-083322P.
28-APR-1998; 98US-083495P.
29-APR-1998; 98US-083496P.
29-APR-1998; 98US-083499P.
29-APR-1998; 98US-083559P.
05-MAY-1998; 98US-084366P.
06-MAY-1998; 98US-084414P.
07-MAY-1998; 98US-084639P.
07-MAY-1998; 98US-084640P.
07-MAY-1998; 98US-084643P.
15-MAY-1998; 98US-085579P.
15-MAY-1998; 98US-085580P.
15-MAY-1998; 98US-085822P.
15-MAY-1998; 98US-085700P.
18-MAY-1998; 98US-086023P.
22-MAY-1998; 98US-086392P.
22-MAY-1998; 98US-086486P.
28-MAY-1998; 98US-087098P.
28-MAY-1998; 98US-087208P.
02-JUN-1998; 98US-087609P.
02-JUN-1998; 98US-087759P.
03-JUN-1998; 98US-087827P.
04-JUN-1998; 98US-088025P.
```


| | | |
|----|--------------|---------------|
| PR | 01-SEP-1998; | 98US-098733P; |
| PR | 02-SEP-1998; | 98US-098803P; |
| PR | 02-SEP-1998; | 98US-098821P; |
| PR | 02-SEP-1998; | 98US-098843P; |
| PR | 09-SEP-1998; | 98US-098602P; |
| PR | 10-SEP-1998; | 98US-099741P; |
| PR | 10-SEP-1998; | 98US-099754P; |
| PR | 10-SEP-1998; | 98US-099763P; |

| Seq. local similarity | Seq. ID | Accession No. | Length | Matches | Mismatches | Indels | Gaps |
|-----------------------|---------|---------------|--------|---------|------------|--------|------|
| 100% | 100% | 100% | 100% | 204 | 0 | 0 | 1 |

| | | | |
|----|-----|---|-----|
| Qy | 1 | GRLLS-----GQPCVCGGTQPCVKYIYFHTSRRLNFEFEAKCAECRRGGQLVSTES | 52 |
| Db | 24 | GRLLSADLDLRGGQPCVCGGTQPCVKYIYFHTSRRLNFEFEAKCAECRRGGQLVSTES | 83 |
| Qy | 53 | EDEQKLEKFIENLLPSDGFWIGLRREKQSNSTACQDLYAWTDGSIISOFRNWYVDEP | 112 |
| Db | 84 | EDEQKLEKFIENLLPSDGFWIGLRREKQSNSTACQDLYAWTDGSIISOFRNWYVDEP | 143 |
| Qy | 113 | SCGSEVCVMYHOPSAAGIGGPFYMFQWDDRCNMKNFIICKYSDEKPAVPSREAEGRET | 172 |
| Db | 144 | SCGSEVCVMYHQPSPAAGIGGPFYMFQWDDRCNMKNFIICKYSDEKPAVPSREAEGRET | 203 |
| Qy | 173 | ELTPVLPEETQBEDAKTTFKESREAAINLAY | 204 |
| Db | 204 | ELTPVLPEETQBEDAKTTFKESREAAINLAY | 235 |

Search completed: December 22, 2003, 16:10:08
Job time : 29.8986 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:04:13 ; Search time 28.7569 Seconds
(without alignments)
1120.478 Million cell updates/sec

Title: US-09-887-855-2_COPY_25_227

Perfect score: 1100

Sequence: 1 RLISGQVCRGGRPCYKV.....EDAKTKFKSREAAALNAY 203

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 1100 | 100.0 | 374 | 21 | AAV93948 |
| 2 | 1100 | 100.0 | 374 | 22 | AAE03651 |
| 3 | 1100 | 100.0 | 374 | 23 | ABG66680 |
| 4 | 1100 | 100.0 | 374 | 23 | ABE90203 |
| 5 | 1100 | 100.0 | 387 | 22 | AAW25796 |
| 6 | 1090 | 99.3 | 374 | 21 | AAV91490 |
| 7 | 1086 | 98.7 | 382 | 20 | AAV13367 |
| 8 | 1086 | 98.7 | 382 | 22 | AAU29033 |
| 9 | 1086 | 98.7 | 382 | 22 | AAE80235 |

| | | | | | | |
|----|-------|------|-----|----|----------|--------------------|
| 10 | 1086 | 98.7 | 382 | 24 | ABU69645 | Novel human secret |
| 11 | 1086 | 98.7 | 382 | 24 | ABU71121 | Human PRO234 prote |
| 12 | 1086 | 98.7 | 382 | 24 | ABU71468 | Human PRO polypept |
| 13 | 1086 | 98.7 | 382 | 24 | ABU71914 | Human secreted/tra |
| 14 | 1086 | 98.7 | 382 | 24 | ABU65578 | Human secreted/tra |
| 15 | 1086 | 98.7 | 382 | 24 | ABU65911 | Novel human secret |
| 16 | 1086 | 98.7 | 382 | 24 | ABU67368 | Human secreted pro |
| 17 | 1086 | 98.7 | 382 | 24 | ABU67415 | Human secreted/tra |
| 18 | 1086 | 98.7 | 382 | 24 | ABU64522 | Human secreted/tra |
| 19 | 1086 | 98.7 | 382 | 24 | ABU65273 | Human PRO polypept |
| 20 | 1086 | 98.7 | 382 | 24 | ABU58409 | Human PRO polypept |
| 21 | 1086 | 98.7 | 382 | 24 | ABU55945 | Human secreted/tra |
| 22 | 1086 | 98.7 | 382 | 24 | ABU56940 | Human PRO polypept |
| 23 | 1086 | 98.7 | 382 | 24 | ABU54370 | Human secreted/tra |
| 24 | 1086 | 98.7 | 382 | 24 | ABU10519 | Human secreted/tra |
| 25 | 561.5 | 51.0 | 273 | 21 | AAE18913 | A novel polypeptid |
| 26 | 561.5 | 51.0 | 273 | 22 | AAU12441 | Human PRO1890 poly |
| 27 | 561.5 | 51.0 | 273 | 22 | AAE73309 | Human C-type lecti |
| 28 | 561.5 | 51.0 | 273 | 22 | AAE87609 | Human PRO1890. Ho |
| 29 | 561.5 | 51.0 | 273 | 23 | ABG95934 | Human secreted/tra |
| 30 | 561.5 | 51.0 | 273 | 23 | ABE95554 | Human angiogenesis |
| 31 | 561.5 | 51.0 | 273 | 23 | ABE84948 | Human PRO1890 prot |
| 32 | 561.5 | 51.0 | 273 | 24 | ABU69084 | Human PRO polypept |
| 33 | 561.5 | 51.0 | 273 | 24 | ABU69107 | Human PRO polypept |
| 34 | 561.5 | 51.0 | 273 | 24 | ABU71589 | Human secreted pol |
| 35 | 561.5 | 51.0 | 273 | 24 | ABU72035 | Novel human secret |
| 36 | 561.5 | 51.0 | 273 | 24 | ABU72192 | Human PRO polypept |
| 37 | 561.5 | 51.0 | 273 | 24 | ABU66839 | Human secreted/tra |
| 38 | 561.5 | 51.0 | 273 | 24 | ABU67115 | Human secreted/tra |
| 39 | 561.5 | 51.0 | 273 | 24 | ABU59920 | Novel secreted and |
| 40 | 534.5 | 48.6 | 274 | 23 | ABF69211 | Human polypeptide |
| 41 | 504.5 | 45.9 | 232 | 22 | ABE94192 | Human protein sequ |
| 42 | 275 | 25.0 | 102 | 22 | ABG51278 | Human liver peptid |
| 43 | 275 | 25.0 | 102 | 22 | ABE21788 | Protein #3787 enco |
| 44 | 275 | 25.0 | 102 | 22 | AAE29941 | Peptide #3978 enco |
| 45 | 225 | 20.5 | 81 | 21 | AAV91643 | Human secreted pro |

ALIGNMENTS

| | |
|----------|--|
| RESULT 1 | |
| AAV93948 | |
| ID | AAV93948 standard; Protein; 374 AA. |
| XX | AC |
| XX | AAV93948; |
| XX | 03-OCT-2000 (first entry) |
| XX | Amino acid sequence of a lectin ss3939 polypeptide. |
| DE | Human; lectin ss3939; chromosome 11; gene therapy. |
| XX | Homo sapiens. |
| OS | |
| XX | Key |
| FH | Location/Qualifiers |
| FT | Peptide |
| FT | 1..21 |
| FT | /note= "signal peptide" |
| FT | Domain |
| FT | 22..227 |
| FT | /note= "extracellular coding region" |
| FT | Domain |
| FT | 228..248 |
| FT | /note= "predicted transmembrane domain" |
| FT | Domain |
| FT | 249..374 |
| FT | /note= "predicted cytoplasmic or intracellular domain" |
| XX | |
| XX | WO200039296-A1. |
| PN | |
| XX | |
| XX | 06-JUL-2000. |
| XX | |
| XX | 22-DEC-1999; |
| XX | 99WO-US30523. |
| XX | |
| XX | 23-DEC-1998; |
| XX | 98US-0113820. |

PA (IMMV) IMMUNEX CORP.
 XX Anderson DA;
 XX WPI; 2000-452394/39.
 DR N-PSDB; AAA57382.
 XX ss3939 nucleic acids, polypeptides and antibodies, useful for
 PT identifying human chromosome 11 and diseases associated with it -
 XX Claim 12; Page 8; 73pp; English.
 XX The present sequence represents a human lectin ss3939 polypeptide. The
 CC polynucleotide sequence is a source of probes, which may be used
 CC to identify nucleic acids encoding ss3939 proteins, to identify human
 CC chromosome number 11, to map genes on human chromosome 11, to
 CC identify diseases associated with chromosome 11, as single-stranded
 CC sense or antisense oligonucleotides to inhibit expression of
 CC polypeptides encoded by the ss3939 gene, and for gene therapy. The
 CC ss3939 polypeptides may be useful for developing treatments for
 CC diseases (none specified) associated with defective or insufficient
 CC amounts of the polypeptides. The antibodies may be useful for
 CC detecting the presence of ss3939 polypeptides.
 XX Sequence 374 AA;
 SQ

Query Match 100.0%; Score 1100; DB 21; Length 374;
 Best Local Similarity 100.0%; Pred. No. 1e-103;
 Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLLSGQVPCRGGTQPCYKVIYFHDTSRLNPFEEAKCRDGGQLVSISEDEQKLEK 60
 Db 25 RLLSGQVPCRGGTQPCYKVIYFHDTSRLNPFEEAKCRDGGQLVSISEDEQKLEK 84
 QY 61 FIENLLPSDGDWIGLRRREKQNSTACODLYAWTDGSIQFRNRYVDEPSCGVV 120
 Db 85 FIENLLPSDGDWIGLRRREKQNSTACODLYAWTDGSIQFRNRYVDEPSCGVV 144
 QY 121 MYHQPSAPAGIGPGYFQWNDRCNNKNNFICKYSDKPAVPSREAGSETELTTPVLPE 180
 Db 145 MYHQPSAPAGIGPGYFQWNDRCNNKNNFICKYSDKPAVPSREAGSETELTTPVLPE 204
 QY 181 ETQEDAKTKFESREAAALNLAY 203
 Db 205 ETQEDAKTKFESREAAALNLAY 227

RESULT 2
 ID AAE03651 standard; Protein; 374 AA.
 AC AAE03651;
 XX 06-AUG-2001 (first entry)
 DT Human extracellular matrix and cell adhesion molecule-15 (XMAD-15).
 XX Human; extracellular matrix and cell adhesion molecule; XMAD;
 KW gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
 KW Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
 KW sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma;
 KW inflammatory disorder; acquired immune deficiency syndrome; AIDS;
 KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
 KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
 KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
 KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
 KW infection; cell proliferative disorder; actinic keratosis; myeloma;
 KW arteriosclerosis; neutropenic; anticonvulsant; antithyroid; nephrotropic;
 KW neuroprotective; dermatological.
 XX Homo sapiens.
 OS
 XX FH Key Location/Qualifiers

FT Peptide 1..24 /label= Signal_peptide
 FT Protein 25..374 /note= "Mature human extracellular matrix and cell
 FT adhesion molecule (XMAD)"
 FT Domain 46..63 /note= "C-type lectin domain"
 FT Domain 163..176 /note= "C-type lectin domain"
 FT Domain 224..247 /note= "Transmembrane motif"
 FT Domain 328..348 /note= "Transmembrane motif"
 XX WO200142285-A2.
 XX 14-JUN-2001.
 XX 05-DEC-2000; 2000WO-US32990.
 XX 10-DEC-1999; 99US-0172852.
 XX 16-DEC-1999; 99US-0172354.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
 PI Baughn MR, Lu DAM, Shah P, Au-Young J;
 XX WPI; 2001-381632/40.
 DR N-PSDB; AAD08059.
 XX New human extracellular matrix and cell adhesion molecules and
 PT polynucleotide sequences encoding them, useful for diagnosis,
 PT prevention, treatment of genetic, autoimmune and cell proliferative
 PT disorders
 XX Claim 1; Page 108-109; 135pp; English.
 XX The present sequence is a human extracellular matrix and cell
 CC adhesion molecule (XMAD). The XMAD is used for screening a compound for
 CC effectiveness as an agonist or antagonist of XMAD. The identified agonist
 CC or antagonist are used for treating a disease or condition associated
 CC with decreased or increased expression of functional XMAD. The
 CC polynucleotides encoding XMAD are useful in somatic or germline gene
 CC therapy to correct a genetic deficiency, to express a conditionally
 CC lethal gene product and to express a protein which affords protection
 CC against intracellular parasites and also for diagnosis of disorders
 CC associated with expression of XMAD. They are also used for generating
 CC hybridisation probes useful in mapping the naturally occurring genomic
 CC sequences and to create knock in humanised animals (pigs) or transgenic
 CC animals (mice or rats) to model human diseases. Oligonucleotide or longer
 CC fragments derived from the polynucleotide sequences may be used as
 CC elements on a microarray. Antibodies which specifically bind XMAD may be
 CC used for the diagnosis of disorders associated with the expression of
 CC XMAD, or in assays to monitor patients being treated with XMAD. Diseases
 CC diagnosed, prevented or treated include genetic disorders such as
 CC adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
 CC disease, myotonic dystrophy, sickle cell anaemia, thalassaemia,
 CC autoimmune/inflammatory disorders such as acquired immune deficiency
 CC syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
 CC atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,
 CC glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
 CC osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis, and
 CC bacterial, fungal, parasitic, protozoal and helminthic infections, and
 CC cell proliferative disorders such as actinic keratosis, arteriosclerosis
 CC and cancer including breast, bladder, bone marrow, brain and uterus
 CC cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.
 XX Sequence 374 AA;
 SQ

Query Match 100.0%; Score 1100; DB 22; Length 374;
 Best Local Similarity 100.0%; Pred. No. 1e-103;
 Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLLSGQVCRGGTQPCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSIIESEDEQKLIK 60
 Db 25 RLLSGQVCRGGTQPCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSIIESEDEQKLIK 84
 Qy 61 FIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNWTVDPSGSEVCV 120
 Db 85 FIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNWTVDPSGSEVCV 144
 Qy 121 MYHOPSAPAGIGGYPYFQWDDRCNMKNFICKYSDEKPAVPSREARAGEETELTPVLPE 180
 Db 145 MYHOPSAPAGIGGYPYFQWDDRCNMKNFICKYSDEKPAVPSREARAGEETELTPVLPE 204
 Qy 181 ETQEDAKTKFKESREAAALNLAY 203
 Db 205 ETQEDAKTKFKESREAAALNLAY 227

RESULT 3

ABG66680
 ID ABG66680 standard; Protein; 374 AA.

AC ABG66680;

DT 30-AUG-2002 (first entry)

DE Human novel polypeptide #15.

Human; inflammatory condition; shock; sepsis; immune response;
 cancer; wound healing; central nervous system disease; haematopoiesis;
 peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
 myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
 cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
 bone degenerative disorder; periodontal disease; reperfusion injury;
 lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
 allergic condition; thrombolytic; thrombosis; coagulation disorder;
 fungal infection.

OS Homo sapiens.

PN WO200244340-A2.

PD 06-JUN-2002.

PF 30-NOV-2001; 2001WO-US47004.

PR 30-NOV-2000; 2000US-0028952.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
 Yamazaki V, Ujwal ML, Drmanac RT;

PP WPI; 2002-508509/54.

DR N-PSDB; ABK94904.

PT Novel nucleic acids and polypeptides for diagnosis, treatment of
 inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
 disorders, cancer and promoting wound healing -

PS Claim 10; Page 579-580; 672pp; English.

XX The invention relates to human novel polynucleotides and associated
 CC polypeptides. The polynucleotides and polypeptides are useful for
 CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
 CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
 CC and cancer and for promoting wound healing. The sequences are used to
 CC induce the proliferation of neural cells and regeneration of nerve and
 CC brain tissue, and are useful for the treatment of central and peripheral
 CC nervous system diseases and neuropathies, such as Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease and amyotrophic lateral
 CC sclerosis. The sequences are involved in chemotactic or chemokinetic
 CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid

cell disorders and platelet disorders such as thrombocytopenia,
 CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
 CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
 CC disease. The sequences of the invention are also useful for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues, immune deficiencies and disorders
 CC including severe combined immunodeficiency (SCID), bacterial or fungal
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
 CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic
 CC and coagulation disorders. Sequences ABG66666-ABG666758 represent human
 CC novel polypeptides of the invention.

XX Sequence 374 AA;

Query Match 100.0%; Score 1100; DB 23; Length 374;

Best Local Similarity 100.0%; Pred. No. 1e-103;

Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLLSGQVCRGGTQPCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSIIESEDEQKLIK 60

Db 25 RLLSGQVCRGGTQPCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSIIESEDEQKLIK 84

Qy 61 FIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNWTVDPSGSEVCV 120

Db 85 FIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNWTVDPSGSEVCV 144

Qy 121 MYHOPSAPAGIGGYPYFQWDDRCNMKNFICKYSDEKPAVPSREARAGEETELTPVLPE 180

Db 145 MYHOPSAPAGIGGYPYFQWDDRCNMKNFICKYSDEKPAVPSREARAGEETELTPVLPE 204

Qy 181 ETQEDAKTKFKESREAAALNLAY 203

Db 205 ETQEDAKTKFKESREAAALNLAY 227

RESULT 4

ABB90203

ID ABB90203 standard; Protein; 374 AA.

AC ABB90203;

DT 24-MAY-2002 (first entry)

DE Human polypeptide SEQ ID NO 2579.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.

OS Homo sapiens.

XX WO200190304-A2.

XX 29-NOV-2001.

XX 18-MAY-2001; 2001WO-US16450.

XX 19-MAY-2000; 2000US-205515P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Birse CE, Rosen CA;

XX WPI; 2002-122018/16.

XX N-PSDB; ABL90612.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -

```

XX PS Claim 11; SEQ ID NO 2579; 2081pp + Sequence Listing; English.
XX CC
XX CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ Sequence 374 AA;
Query Match 100.0%; Score 1100; DB 23; Length 374;
Best Local Similarity 100.0%; Pred. No. 1e-103;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RLLSQPVCRCGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQGVLSIESDEQKLEK 60
DB 25 RLLSQPVCRCGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQGVLSIESDEQKLEK 84
QY 61 FIENLLPSDGFWIGLRRREKQSNSTACQDLVYATDGSISQFRNMYVDPSGSEVCV 120
DB 85 FIENLLPSDGFWIGLRRREKQSNSTACQDLVYATDGSISQFRNMYVDPSGSEVCV 144
QY 121 MYHQPAPAGIGGPPYMFQWDDRCNMKNPFICKYSDKPAVPSRAEGEETLTTPVLPE 180
DB 145 MYHQPAPAGIGGPPYMFQWDDRCNMKNPFICKYSDKPAVPSRAEGEETLTTPVLPE 204
QY 181 ETQEDAKTKFKESREAAALNLAY 203
DB 205 ETQEDAKTKFKESREAAALNLAY 227
RESULT 5
AAM25796
ID AAM25796 standard; Protein; 387 AA.
XX AC AAM25796;
XX DT
XX DT 16-OCT-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:1311.
XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiant; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.
XX OS Homo sapiens.
XX

```

```

PN WO200153455-A2.
XX PD 26-JUL-2001.
XX XX
XX 22-DEC-2000; 2000WO-US35017.
XX 23-DEC-1999; 99US-0471275.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Dermanan RT;
XX WPI: 2001-457603/49.
XX N-PSDB; AAH99737.
XX Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX Claim 20; Page 272; 1217pp; English.
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
AAM25963. The proteins can have activities based on the tissues and
cells they are expressed in, such as: antiinflammatory; antirheumatic;
antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
central nervous system; virucide; anti-HIV; fungicide; antimutagen;
cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
antiparkinsonian; and immunostimulant. The proteins and polynucleotides
encoding them can be used in gene therapy, antisense therapy and vaccine
production. The proteins and polynucleotides are useful for screening for
agonists or antagonists of a protein and for the treatment and diagnosis
of disorders associated with the activity of a protein e.g. inflammation,
rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
infections, autoimmunity, genetic diseases, haematopoietic disorders,
anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
osteoporosis, severe combined immunodeficiency, eczema, allergic
rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
Alzheimer's disease, Parkinson's disease, neurodegenerative and
neurological disorders.
XX SQ Sequence 387 AA;
Query Match 100.0%; Score 1100; DB 22; Length 387;
Best Local Similarity 100.0%; Pred. No. 1.e-103;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RLLSQPVCRCGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQGVLSIESDEQKLEK 60
DB 38 RLLSQPVCRCGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQGVLSIESDEQKLEK 97
QY 61 FIENLLPSDGFWIGLRRREKQSNSTACQDLVYATDGSISQFRNMYVDPSGSEVCV 120
DB 98 FIENLLPSDGFWIGLRRREKQSNSTACQDLVYATDGSISQFRNMYVDPSGSEVCV 157
QY 121 MYHQPAPAGIGGPPYMFQWDDRCNMKNPFICKYSDKPAVPSRAEGEETLTTPVLPE 180
DB 158 MYHQPAPAGIGGPPYMFQWDDRCNMKNPFICKYSDKPAVPSRAEGEETLTTPVLPE 217
QY 181 ETQEDAKTKFKESREAAALNLAY 203
DB 218 ETQEDAKTKFKESREAAALNLAY 240
RESULT 6
AAY91490
ID AAY91490 standard; Protein; 374 AA.
XX AC AAY91490;
XX

```

29-JUN-2000 (first entry)
 Human secreted protein sequence encoded by gene 40 SEQ ID NO:163.
 Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 antiHIV; antineoplastic; neuroprotective; antidiabetic;
 osteoporosis; antiparasitic; antibacterial; antidiabetic; antiasthma;
 antiparasitic; cardiant; gene therapy; cancer; neurological disorder;
 immune disease; inflammation; blood disorder; tumour.
 Homo sapiens.
 WO200006698-A1.
 10-FEB-2000.
 29-JUL-1999; 99WO-US17130.
 30-JUL-1998; 98US-0094657.
 05-AUG-1998; 98US-0095486.
 06-AUG-1998; 98US-0095454.
 06-AUG-1998; 98US-0095455.
 12-AUG-1998; 98US-0096319.
 (HUMA-) HUMAN GENOME SCI INC.
 Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
 Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
 Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
 WPI; 2000-195282/17.
 N-PSDB; AAA26385.
 New isolated human genes and the secreted polypeptides they encode,
 useful for diagnosis and treatment of e.g. cancers, neurological
 disorders, immune diseases, inflammation or blood disorders -
 Claim 11; Page 483-484; 634pp; English.
 The polynucleotide sequences given in AAA26346 to AAA26458 encode the
 human secreted proteins given in AAY91451 to AAY91691. The human secreted
 proteins can have activities based on the tissues and cells they are
 expressed in. Examples of the activities are: cytostatic;
 immunosuppressive; antiHIV; antineoplastic; neuroprotective;
 antidiabetic; osteoporosis; antiparasitic; antibacterial; antidiabetic;
 antiasthma; antiparasitic; and cardiant. The polynucleotides and their
 corresponding secreted proteins are useful for preventing, treating or
 ameliorating medical conditions, e.g. by protein or gene therapy. Also
 pathological conditions can be diagnosed by determining the amount of the
 proteins in a sample or by determining the presence of mutations in the
 polynucleotides. Specific uses are described for each of the
 polynucleotides, based on which tissues they are most highly expressed
 in, and include developing products for the diagnosis or treatment of
 cancer, tumours, neurodegenerative disorders, developmental abnormalities
 and foetal deficiencies, blood disorders, diseases of the immune system,
 autoimmune diseases, hepatic and renal disease, inflammation,
 allergies, Alzheimer's and behavioural disorders, schizophrenia,
 osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 cardiovascular disorders, reproductive disorders, gastrointestinal
 disorders, respiratory disorders and metabolic disorders. The proteins
 or polynucleotides can also be used as food additives or preservatives.
 The proteins are also useful for identifying their binding partners.
 AAA26337 to AAA26345 and AAY91450 are sequences used in the
 exemplification of the present invention.
 Sequence 374 AA;
 Query Match 99.1%; Score 1090; DB 21; Length 374;
 Best Local Similarity 99.0%; Pred. No. 1.1e-102;
 Matches 201; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 1 RLLSGQPVCRGGTQPCVKYIYFHTSRLNFEAKEACRDGGQLVSEDEQKLIEX 60

25 RLLSGQPVCRGGTQPCVKYIYFHTSRLNFEAKEACRDGGQLVSEDEQKLIEX 84
 61 FIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPSCGSEVCV 120
 85 FIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPSCGSEVCV 144
 121 MYHOPSAPAGIGGPFYMFQWDDRCNKNFICKYSDEKPAVPSREAGEETEELTPVLPE 180
 145 MYHOPSAPAGIGGPFYMFQWDDRCNKNFICKYSDEKPAVPSREAGEETEELTPVLPE 204
 181 ETOEDAKTFFKESREAAALNLAY 203
 205 ETOEDAKTFFKESREAAALNLAY 227
 RESULT 7
 AAY13367
 ID AAY13367 standard; Protein; 382 AA.
 AC AAY13367;
 XX
 DT 25-JUN-1999 (first entry)
 DE
 DE Amino acid sequence of protein PRO234.
 KW Secreted protein; transmembrane protein; human; enterocolitis;
 KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
 KW congenital microvillus atrophy; skin disease; cell growth;
 KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
 KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
 KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
 KW anti-thrombotic; wound healing; tissue repair.
 OS Homo sapiens.
 XX
 XX WO9914328-A2.
 XX
 PD 25-MAR-1999.
 XX
 PF 16-SEP-1998; 98WO-US19330.
 XX
 PR 25-NOV-1997; 97US-0066840.
 PR 17-SEP-1997; 97US-0059113.
 PR 17-SEP-1997; 97US-0059115.
 PR 17-SEP-1997; 97US-0059117.
 PR 17-SEP-1997; 97US-0059119.
 PR 17-SEP-1997; 97US-0059121.
 PR 17-SEP-1997; 97US-0059122.
 PR 17-SEP-1997; 97US-0059184.
 PR 18-SEP-1997; 97US-0059263.
 PR 18-SEP-1997; 97US-0059266.
 PR 15-OCT-1997; 97US-0062125.
 PR 17-OCT-1997; 97US-0062285.
 PR 17-OCT-1997; 97US-0062287.
 PR 21-OCT-1997; 97US-0063486.
 PR 24-OCT-1997; 97US-0062814.
 PR 24-OCT-1997; 97US-0062816.
 PR 24-OCT-1997; 97US-0063045.
 PR 24-OCT-1997; 97US-0063120.
 PR 24-OCT-1997; 97US-0063121.
 PR 24-OCT-1997; 97US-0063127.
 PR 24-OCT-1997; 97US-0063128.
 PR 27-OCT-1997; 97US-0063329.
 PR 27-OCT-1997; 97US-0063327.
 PR 28-OCT-1997; 97US-0063541.
 PR 28-OCT-1997; 97US-0063542.
 PR 28-OCT-1997; 97US-0063544.
 PR 28-OCT-1997; 97US-0063549.
 PR 28-OCT-1997; 97US-0063550.
 PR 29-OCT-1997; 97US-0063564.
 PR 29-OCT-1997; 97US-0063435.
 PR 29-OCT-1997; 97US-0063704.

PR 29-OCT-1997; 97US-0063732.
 PR 29-OCT-1997; 97US-0063738.
 PR 29-OCT-1997; 97US-0063734.
 PR 29-OCT-1997; 97US-0064215.
 PR 29-OCT-1997; 97US-0063735.
 PR 31-OCT-1997; 97US-0063870.
 PR 31-OCT-1997; 97US-0064103.
 PR 03-NOV-1997; 97US-0064248.
 PR 07-NOV-1997; 97US-0064809.
 PR 12-NOV-1997; 97US-0065186.
 PR 17-NOV-1997; 97US-0065846.
 PR 18-NOV-1997; 97US-0065693.
 PR 21-NOV-1997; 97US-0066120.
 PR 21-NOV-1997; 97US-0066364.
 PR 24-NOV-1997; 97US-0066772.
 PR 24-NOV-1997; 97US-0066466.
 PR 24-NOV-1997; 97US-0066770.
 PR 24-NOV-1997; 97US-0066511.
 PR 24-NOV-1997; 97US-0066453.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 XX WPI; 1999-229533/19.
 DR N-PSDB; AAX52238.
 DR
 XX New isolated human genes and polypeptides used in, e.g. treatment of
 PT gastrointestinal ulceration
 PT
 XX Claim 12; Fig 50; 320pp; English.
 XX
 CC AAY13344-403 represent secreted and transmembrane human proteins.
 CC The cDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO11 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to
 CC growth or survival of nerve cells including Parkinson's disease,
 CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
 CC for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata; PRO269 can be used as an
 CC anti-thrombotic agent; PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair; PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.
 XX
 XX Sequence 382 AA;
 SQ
 Query Match 98.7%; Score 1086; DB 20; Length 382;
 Best Local Similarity 96.2%; Pred. No. 2.8e-102;
 Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 RLLSASDLDLRGGQPCVCGTQPCVKVIYFHDTSRLNFEAKEACRRDGGQLVSTSE 52
 DB 25 RLLSASDLDLRGGQPCVCGTQPCVKVIYFHDTSRLNFEAKEACRRDGGQLVSTSE 84
 QY 53 DEQKLEKFTENLLPSDGFHWIGLRREKQKSNSTACQDLYAWTDCSISQFRNYYDEPS 112
 DB 85 DEQKLEKFTENLLPSDGFHWIGLRREKQKSNSTACQDLYAWTDCSISQFRNYYDEPS 144
 QY 113 CGSEVCVMYHQPAPAGIGGPTFMQWDDRCNMKNFICKYDEKPAVPSRAEGEETE 172
 DB 145 CGSEVCVMYHQPAPAGIGGPTFMQWDDRCNMKNFICKYDEKPAVPSRAEGEETE 204
 QY 173 LITPVLPEETQBEDAKKTFKESREALNLAY 203

Db 205 LITPVLPEETQBEDAKKTFKESREALNLAY 235
 RESULT 8
 AAU29033
 ID AAU29033 standard; Protein; 382 AA.
 XX
 AC AAU29033;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human PRO polypeptide sequence #10.
 XX
 KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; INF-alpha;
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200168848-A2.
 XX
 PD 20-SEP-2001.
 XX
 PP 28-FEB-2001; 2001WO-US06520.
 XX
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 03-MAR-2000; 2000US-187202P.
 PR 06-MAR-2000; 2000US-186968P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 21-MAR-2000; 2000US-190828P.
 PR 21-MAR-2000; 2000US-191007P.
 PR 21-MAR-2000; 2000US-191048P.
 PR 21-MAR-2000; 2000US-191314P.
 PR 28-MAR-2000; 2000US-192655P.
 PR 28-MAR-2000; 2000US-193032P.
 PR 29-MAR-2000; 2000US-193053P.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 04-APR-2000; 2000US-194449P.
 PR 04-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 25-APR-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 22-AUG-2000; 2000US-064848.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 (GETH) GENENTECH INC.
 PA
 XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX WPI; 2001-602746/68.
 DR

DR N-PSDB; AAS45934.

XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the

PT presence of tumours, such as prostate and breast tumours, in mammals and

PT to screen for modulators of the compounds -

XX

PS Claim 11; Fig 20; 774pp; English.

XX

XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.

CC The PRO polypeptides and their associated nucleic acids can be used to

CC detect the presence of a tumour in a mammal by comparing the level of

CC expression of a PRO polypeptide in a test sample of cells from the animal

CC and a control sample of normal cells, whereby a higher level of

CC expression in the test sample indicates the presence of a tumour in the

CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats

CC and rabbits but are preferably human. The polypeptides can be used to

CC stimulate tumour necrosis factor (TNF) alpha release from human blood,

CC when contacted with it. A specific polypeptide can be used to stimulate

CC the proliferation or differentiation of chondrocyte cells. The PRO

CC proteins can be used to determine the presence of tumours and also

CC susceptibility to tumour development, particularly adrenal, lung, colon,

CC breast, prostate, rectal, cervical, or liver tumours, in mammalian

CC subjects. The oligonucleotide probes specific for the PRO nucleic acids

CC can be used for genetic analysis of individuals with genetic disorders.

XX

SQ Sequence 382 AA;

Query Match 98.7%; Score 1086; DB 22; Length 382;

Best Local Similarity 96.2%; Pred. No. 2.8e-102;

Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 RLJS-----GQVCRGQTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSTSE 52

Db 25 RLLSASDLRLGGQPVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSTSE 84

Qy 53 DEOKLIEKFLENLPSDGFWIGLRREEKQSNSTACODLYAWTDGSIQFRNMYVDEPS 112

Db 85 DEOKLIEKFLENLPSDGFWIGLRREEKQSNSTACODLYAWTDGSIQFRNMYVDEPS 144

Qy 113 CGSEVCVMYHQSPAPAGIGPYMFQWDDRCNKNFNICKYSDEKPAVPSREAGBETE 172

Db 145 CGSEVCVMYHQSPAPAGIGPYMFQWDDRCNKNFNICKYSDEKPAVPSREAGBETE 204

Qy 173 LTTPLVPEETQEDAKTKFKESREAAALNLAY 203

Db 205 LTTPLVPEETQEDAKTKFKESREAAALNLAY 235

RESULT 9

AAB80235

ID AAB80235 standard; Protein; 382 AA.

XX

AC AAB80235;

XX

XX 24-APR-2001 (first entry)

XX

DE Human PRO234 protein.

XX

XX Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;

KW antiparkinsonian nootropic; neuroprotective; vulnery; cardiant;

KW antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;

KW antiarthritic; antifertility; antidiabetic; antiviral; diabetes;

KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;

KW ischaemia; inflammation.

XX

OS Homo sapiens.

XX

XX WO200104311-A1.

XX

XX 18-JAN-2001.

XX

XX 22-FEB-2000; 2000WO-US04414.

XX

PR 07-JUL-1999; 99US-0143048.

PR 26-JUL-1999; 99US-0145698.

PR 28-JUL-1999; 99US-0146222.

PR 08-SEP-1999; 99WO-US20594.

PR 13-SEP-1999; 99WO-US20944.

PR 15-SEP-1999; 99WO-US21090.

PR 15-SEP-1999; 99WO-US21547.

PR 05-OCT-1999; 99WO-US23089.

PR 29-NOV-1999; 99WO-US28214.

PR 30-NOV-1999; 99WO-US28313.

PR 16-DEC-1999; 99WO-US30095.

PR 20-DEC-1999; 99WO-US30911.

PR 20-DEC-1999; 99WO-US30999.

PR 05-JAN-2000; 99WO-US00219.

XX (GETH) GENENTECH INC.

XX

XX Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;

PI Filvaroff E, Fong S, Gao W, Gerber H, Gertitsen ME, Goddard A;

PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin IJ;

PI Mather JP, Pan J, Paooni NF, Roy MA, Stewart TA, Tumas D;

PI Williams PM, Wood WI;

XX WPI; 2001-081051/09.

DR N-PSDB; AAF72396.

XX

PT Sixty one nucleic acids encoding PRO polypeptides which are useful in

PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung

PT squamous cell carcinoma) and neurodegenerative diseases (e.g.

PT Alzheimer's disease) -

XX

Claim 1; Fig 50; 393pp; English.

XX

XX The present sequence is one of sixty one novel secreted and

CC transmembrane PRO polypeptides. The PRO polypeptides are

CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung

CC squamous cell carcinoma), gastrointestinal disorders (e.g.

CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,

CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.

CC endometrial bleeding, angiogenesis, ischaemias such as coronary

CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,

CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and

CC diabetes and retinal disorders such as retinitis pigmentosum.

CC The PRO nucleic acids have applications in molecular biology, including

CC use as hybridization probes, and in chromosome and gene mapping.

XX

SQ Sequence 382 AA;

Query Match 98.7%; Score 1086; DB 22; Length 382;

Best Local Similarity 96.2%; Pred. No. 2.8e-102;

Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 RLJS-----GQVCRGQTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSTSE 52

Db 25 RLLSASDLRLGGQPVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSTSE 84

Qy 53 DEOKLIEKFLENLPSDGFWIGLRREEKQSNSTACODLYAWTDGSIQFRNMYVDEPS 112

Db 85 DEOKLIEKFLENLPSDGFWIGLRREEKQSNSTACODLYAWTDGSIQFRNMYVDEPS 144

Qy 113 CGSEVCVMYHQSPAPAGIGPYMFQWDDRCNKNFNICKYSDEKPAVPSREAGBETE 172

Db 145 CGSEVCVMYHQSPAPAGIGPYMFQWDDRCNKNFNICKYSDEKPAVPSREAGBETE 204

Qy 173 LTTPLVPEETQEDAKTKFKESREAAALNLAY 203

Db 205 LTTPLVPEETQEDAKTKFKESREAAALNLAY 235

RESULT 10

ABU69645

ID ABU69645 standard; Protein; 382 AA.

XX

| | | | | |
|----|---|----|--------------|----------------|
| AC | ABU69645; | PR | 27-OCT-1997; | 97US-063329P. |
| XX | | PR | 28-OCT-1997; | 97US-063541P. |
| DT | | PR | 28-OCT-1997; | 97US-063542P. |
| XX | | PR | 28-OCT-1997; | 97US-063544P. |
| DE | 05-JUN-2003 (first entry) | PR | 28-OCT-1997; | 97US-063549P. |
| XX | Novel human secreted and transmembrane protein PRO234. | PR | 28-OCT-1997; | 97US-063550P. |
| XX | Human; secreted and transmembrane protein; gene therapy; psoriasis; | PR | 28-OCT-1997; | 97US-063564P. |
| KW | enterocolitis; gastrointestinal ulceration; skin disease; | PR | 28-OCT-1997; | 97US-063435P. |
| KW | keratinocyte differentiation; epithelial cancer; Alzheimer's disease; | PR | 28-OCT-1997; | 97US-0633732P. |
| KW | squamous cell carcinoma; Parkinson's disease; inflammatory disease; | PR | 28-OCT-1997; | 97US-063734P. |
| KW | amyotrophic lateral sclerosis; rheumatoid arthritis; asthma; | PR | 28-OCT-1997; | 97US-063735P. |
| KW | multiple sclerosis; organ failure; atherosclerosis; cardiac injury; | PR | 28-OCT-1997; | 97US-063738P. |
| KW | infertility; birth defect; premature aging; AIDS; cancer; | PR | 28-OCT-1997; | 97US-064215P. |
| KW | diabetic complication; wound repair; tissue re-growth. | PR | 28-OCT-1997; | 97US-063870P. |
| XX | | PR | 28-OCT-1997; | 97US-064103P. |
| OS | Homo sapiens. | PR | 28-OCT-1997; | 97US-064248P. |
| XX | | PR | 28-OCT-1997; | 97US-064809P. |
| PN | US2003017463-A1. | PR | 28-OCT-1997; | 97US-065186P. |
| XX | | PR | 28-OCT-1997; | 97US-065846P. |
| XX | 23-JAN-2003. | PR | 28-OCT-1997; | 97US-065893P. |
| PD | | PR | 28-OCT-1997; | 97US-066120P. |
| XX | | PR | 28-OCT-1997; | 97US-066364P. |
| PF | 11-JUL-2001; 2001US-0903640. | PR | 28-OCT-1997; | 97US-066453P. |
| XX | | PR | 28-OCT-1997; | 97US-066511P. |
| PR | 10-SEP-1998; | PR | 28-OCT-1997; | 97US-066770P. |
| PR | 14-SEP-1998; | PR | 28-OCT-1997; | 97US-066772P. |
| PR | 16-SEP-1998; | PR | 28-OCT-1997; | 97US-066840P. |
| PR | 17-SEP-1998; | PR | 28-OCT-1997; | 97US-066842P. |
| PR | 01-DEC-1998; | PR | 28-OCT-1997; | 97US-066846P. |
| PR | 08-SEP-1999; | PR | 28-OCT-1997; | 97US-066848P. |
| PR | 13-SEP-1999; | PR | 28-OCT-1997; | 97US-066850P. |
| PR | 15-SEP-1999; | PR | 28-OCT-1997; | 97US-066852P. |
| PR | 15-SEP-1999; | PR | 28-OCT-1997; | 97US-066854P. |
| PR | 05-OCT-1999; | PR | 28-OCT-1997; | 97US-066856P. |
| PR | 29-NOV-1999; | PR | 28-OCT-1997; | 97US-066858P. |
| PR | 30-NOV-1999; | PR | 28-OCT-1997; | 97US-066860P. |
| PR | 01-DEC-1999; | PR | 28-OCT-1997; | 97US-066862P. |
| PR | 02-DEC-1999; | PR | 28-OCT-1997; | 97US-066864P. |
| PR | 16-DEC-1999; | PR | 28-OCT-1997; | 97US-066866P. |
| PR | 20-DEC-1999; | PR | 28-OCT-1997; | 97US-066868P. |
| PR | 20-DEC-1999; | PR | 28-OCT-1997; | 97US-066870P. |
| PR | 05-JAN-2000; | PR | 28-OCT-1997; | 97US-066872P. |
| PR | 11-FEB-2000; | PR | 28-OCT-1997; | 97US-066874P. |
| PR | 22-FEB-2000; | PR | 28-OCT-1997; | 97US-066876P. |
| PR | 24-FEB-2000; | PR | 28-OCT-1997; | 97US-066878P. |
| PR | 02-MAR-2000; | PR | 28-OCT-1997; | 97US-066880P. |
| PR | 20-MAR-2000; | PR | 28-OCT-1997; | 97US-066882P. |
| PR | 30-MAR-2000; | PR | 28-OCT-1997; | 97US-066884P. |
| PR | 22-MAY-2000; | PR | 28-OCT-1997; | 97US-066886P. |
| PR | 02-JUN-2000; | PR | 28-OCT-1997; | 97US-066888P. |
| PR | 28-JUL-2000; | PR | 28-OCT-1997; | 97US-066890P. |
| PR | 24-AUG-2000; | PR | 28-OCT-1997; | 97US-066892P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066894P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066896P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066898P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066900P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066902P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066904P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066906P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066908P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066910P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066912P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066914P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066916P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066918P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066920P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066922P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066924P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066926P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066928P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066930P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066932P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066934P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066936P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066938P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066940P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066942P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066944P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066946P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066948P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066950P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066952P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066954P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066956P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066958P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066960P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066962P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066964P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066966P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066968P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066970P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066972P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066974P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066976P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066978P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066980P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066982P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066984P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066986P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066988P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066990P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066992P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066994P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066996P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-066998P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067000P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067002P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067004P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067006P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067008P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067010P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067012P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067014P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067016P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067018P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067020P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067022P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067024P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067026P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067028P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067030P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067032P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067034P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067036P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067038P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067040P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067042P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067044P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067046P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067048P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067050P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067052P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067054P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067056P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067058P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067060P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067062P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067064P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067066P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067068P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067070P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067072P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067074P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067076P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067078P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067080P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067082P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067084P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067086P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067088P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067090P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067092P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067094P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067096P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067098P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067100P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067102P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067104P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067106P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067108P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067110P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067112P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067114P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067116P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067118P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067120P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067122P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067124P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067126P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067128P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067130P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067132P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067134P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067136P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067138P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067140P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067142P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067144P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067146P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067148P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067150P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067152P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067154P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067156P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067158P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067160P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067162P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067164P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067166P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067168P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067170P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067172P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067174P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067176P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067178P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067180P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-067182P. |
| PR | 17-SEP-1997; | PR | 28-OCT-1997; | 97US-0671 |

CC protein electrophoresis purposes. The anti-PRO antibodies may be used
 CC in diagnostic assays for PRO, or for the affinity purification of PRO
 CC from recombinant cell culture or natural sources. This is the amino
 CC acid sequence of a novel human PRO polypeptide.

XX Sequence 382 AA;

Query Match 98.7%; Score 1086; DB 24; Length 382;

Best Local Similarity 96.2%; Pred. No. 2.8e-102;

Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

| | | | | |
|----|-----|--|--|----|
| Qy | 1 | RLLS----- | GGPVCGGTQRCYKVIYFHDTSRLNFEAKACRRDGGQLVSI | 52 |
| Db | 25 | RLLSASDLDRGGQVPCGGTQRCYKVIYFHDTSRLNFEAKACRRDGGQLVSI | 84 | |
| Qy | 53 | DEQKLEKFIENLPSDGFWIGLRRREKOSNSTACODLYAWTDGSIQFRNMYVDEPS | 112 | |
| Db | 85 | DEQKLEKFIENLPSDGFWIGLRRREKOSNSTACODLYAWTDGSIQFRNMYVDEPS | 144 | |
| Qy | 113 | CGSEVGVVMYHQPSAPAGIGGYPYFQWNNDRCKNNKFNICKYDEKPAVPSREAGEETE | 172 | |
| Db | 145 | CGSEVGVVMYHQPSAPAGIGGYPYFQWNNDRCKNNKFNICKYDEKPAVPSREAGEETE | 204 | |
| Qy | 173 | LTTPVLPETQEDAKTTFKESREAAALNLAY | 203 | |
| Db | 205 | LTTPVLPETQEDAKTTFKESREAAALNLAY | 235 | |

RESULT 11

ABU71121
 ID ABU71121 standard; Protein; 382 AA.

XX AC

XX ABU71121;

XX DT 10-JUN-2003 (first entry)

XX DE Human PRO234 protein.

XX KW Human; PRO; secreted; transmembrane; cytostatic; TNF-alpha; blood;
 KW tumour necrosis factor alpha release; Chondrocyte cell; proliferation;
 KW differentiation; tumour; gene therapy.

XX OS Homo sapiens.

XX PN US2003036143-A1.

XX PD 20-FEB-2003.

XX PF 02-JUL-2002; 2002US-0187600.

XX PR 16-SEP-1998; 98WO-US19330.

PR 07-OCT-1998; 98WO-US21141.

PR 01-DEC-1998; 98WO-US25108.

PR 08-MAR-1999; 99WO-US05028.

PR 14-MAY-1999; 99WO-US10733.

PR 02-JUN-1999; 99WO-US12252.

PR 01-SEP-1999; 99WO-US20111.

PR 15-SEP-1999; 99WO-US21090.

PR 01-DEC-1999; 99WO-US28301.

PR 02-DEC-1999; 99WO-US28551.

PR 30-DEC-1999; 99WO-US31274.

PR 03-JAN-2000; 2000WO-US00219.

PR 18-FEB-2000; 2000WO-US04341.

PR 18-FEB-2000; 2000WO-US04342.

PR 22-FEB-2000; 2000WO-US04414.

PR 24-FEB-2000; 2000WO-US05004.

PR 01-MAR-2000; 2000WO-US05601.

PR 02-MAR-2000; 2000WO-US05841.

PR 15-MAR-2000; 2000WO-US06884.

PR 30-MAR-2000; 2000WO-US08439.

PR 17-MAY-2000; 2000WO-US13705.

PR 22-MAY-2000; 2000WO-US14042.

PR 30-MAY-2000; 2000WO-US14941.

| | | |
|----|--------------|-----------------|
| PR | 02-JUN-2000; | 2000WO-US15264. |
| PR | 28-JUL-2000; | 2000WO-US20710. |
| PR | 24-AUG-2000; | 2000WO-US23328. |
| PR | 08-NOV-2000; | 2000WO-US30952. |
| PR | 01-DEC-2000; | 2000WO-US32678. |
| PR | 20-DEC-2000; | 2000WO-US34956. |
| PR | 28-FEB-2001; | 2001WO-US06520. |
| PR | 01-JUN-2001; | 2001WO-US17800. |
| PR | 20-JUN-2001; | 2001WO-US19692. |
| PR | 29-JUN-2001; | 2001WO-US21066. |
| PR | 09-JUL-2001; | 2001WO-US21735. |
| PR | 29-AUG-2001; | 2001WO-US27099. |
| PR | 18-SEP-1997; | 97US-059263P. |
| PR | 18-SEP-1997; | 97US-059266P. |
| PR | 17-OCT-1997; | 97US-062250P. |
| PR | 21-OCT-1997; | 97US-063486P. |
| PR | 24-OCT-1997; | 97US-063120P. |
| PR | 24-OCT-1997; | 97US-063121P. |
| PR | 28-OCT-1997; | 97US-063540P. |
| PR | 28-OCT-1997; | 97US-063541P. |
| PR | 28-OCT-1997; | 97US-063544P. |
| PR | 28-OCT-1997; | 97US-063564P. |
| PR | 29-OCT-1997; | 97US-063734P. |
| PR | 31-OCT-1997; | 97US-063870P. |
| PR | 31-OCT-1997; | 97US-064103P. |
| PR | 13-NOV-1997; | 97US-065311P. |
| PR | 21-NOV-1997; | 97US-066120P. |
| PR | 24-NOV-1997; | 97US-066468P. |
| PR | 24-NOV-1997; | 97US-066772P. |
| PR | 11-DEC-1997; | 97US-069335P. |
| PR | 12-DEC-1997; | 97US-069425P. |
| PR | 17-DEC-1997; | 97US-069870P. |
| PR | 18-DEC-1997; | 97US-068017P. |
| PR | 10-MAR-1998; | 98US-077450P. |
| PR | 11-MAR-1998; | 98US-077632P. |
| PR | 11-MAR-1998; | 98US-077649P. |
| PR | 20-MAR-1998; | 98US-078886P. |
| PR | 20-MAR-1998; | 98US-078939P. |
| PR | 27-MAR-1998; | 98US-079664P. |
| PR | 27-MAR-1998; | 98US-079786P. |
| PR | 31-MAR-1998; | 98US-080107P. |
| PR | 31-MAR-1998; | 98US-080194P. |
| PR | 01-APR-1998; | 98US-080327P. |
| PR | 01-APR-1998; | 98US-080333P. |
| PR | 08-APR-1998; | 98US-081049P. |
| PR | 08-APR-1998; | 98US-081070P. |
| PR | 09-APR-1998; | 98US-081195P. |
| PR | 15-APR-1998; | 98US-081838P. |
| PR | 21-APR-1998; | 98US-082568P. |
| PR | 21-APR-1998; | 98US-082569P. |
| PR | 22-APR-1998; | 98US-082704P. |
| PR | 22-APR-1998; | 98US-082797P. |
| PR | 28-APR-1998; | 98US-083322P. |
| PR | 29-APR-1998; | 98US-083495P. |
| PR | 29-APR-1998; | 98US-083496P. |
| PR | 29-APR-1998; | 98US-083499P. |
| PR | 29-APR-1998; | 98US-083559P. |
| PR | 05-MAY-1998; | 98US-084366P. |
| PR | 06-MAY-1998; | 98US-084414P. |
| PR | 07-MAY-1998; | 98US-084639P. |
| PR | 07-MAY-1998; | 98US-084643P. |
| PR | 15-MAY-1998; | 98US-085579P. |
| PR | 15-MAY-1998; | 98US-085580P. |
| PR | 15-MAY-1998; | 98US-085582P. |
| PR | 18-MAY-1998; | 98US-085700P. |
| PR | 22-MAY-1998; | 98US-086023P. |
| PR | 22-MAY-1998; | 98US-086392P. |
| PR | 28-MAY-1998; | 98US-086486P. |
| PR | 28-MAY-1998; | 98US-087098P. |
| PR | 02-JUN-1998; | 98US-087208P. |
| PR | 02-JUN-1998; | 98US-087609P. |
| PR | 02-JUN-1998; | 98US-087759P. |

```
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088217P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089030P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.

PR 26-AUG-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.
PR 01-SEP-1998; 98US-098723P.
PR 01-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.
PR 10-SEP-1998; 98US-099754P.
PR 10-SEP-1998; 98US-099763P.
PR 10-SEP-1998; 98US-099812P.

Query Match 98.7%; Score 1086; DB 24; Length 382;
Best Local Similarity 96.2%; Pred. No. 2.8e-102;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 RLLS-----GQVCRGCTGTPCYKVIYFHDTSRLNFEAKEACRRDGGQLVIESE 52
DB 25 RLLSASDLRGQGPVCRGCTGTPCYKVIYFHDTSRLNFEAKEACRRDGGQLVIESE 84
QY 53 DEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPS 112
DB 85 DEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPS 144
QY 113 CGSEVCVMYHQPAPAGIGGPFQWDDRCNMKNFICKYSDEKPAVPSREAGEETE 172
DB 145 CGSEVCVMYHQPAPAGIGGPFQWDDRCNMKNFICKYSDEKPAVPSREAGEETE 204
QY 173 LTPVLPPEETQEDAKTFKESREAAALNLAY 203
DB 205 LTPVLPPEETQEDAKTFKESREAAALNLAY 235

RESULT 12
ABU71468
ID ABU71468 standard; Protein; 382 AA.
XX
AC ABU71468;
XX
DT 10-JUN-2003 (first entry)
XX
DE Human PRO polypeptide #24.
KW Human; secreted and transmembrane protein; PRO polypeptide; cancer;
KW Alzheimer's disease; ischaemia; cytostatic; nootropic; vasotrophic;
KW neuroprotective.
XX
OS Homo sapiens.
XX
PN US2002192659-A1.
XX
PD 19-DEC-2002.
XX
PF 10-JUL-2001; 2001US-0902853.
XX
PR 10-SEP-1998; 98WO-US18824.
PR 14-SEP-1998; 98WO-US19177.
PR 16-SEP-1998; 98WO-US19330.
PR 17-SEP-1998; 98WO-US19437.
PR 01-DEC-1998; 98WO-US25108.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 05-JAN-2000; 2000WO-US00219.
PR 11-FEB-2000; 2000WO-US03565.
```

```

PR 22-FEB-2000; 2000WO-US04414.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 17-SEP-1997; 97US-059113P.
PR 17-SEP-1997; 97US-059113P.
PR 17-SEP-1997; 97US-059113P.
PR 18-SEP-1997; 97US-059266P.
PR 15-OCT-1997; 97US-062125P.
PR 17-OCT-1997; 97US-062285P.
PR 17-OCT-1997; 97US-062287P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-062814P.
PR 24-OCT-1997; 97US-062816P.
XX (GETH ) GENENTECH INC.
XX
XX Aekhenazi A, Botstein D, Deenoyers L, Eaton DL, Ferrara N;
PI Filvaroff E, Fong S, Gao W, Garber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kijavini IG;
PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
PI Williams PM, Wood WI;
XX WPI: 2003-361832/34.
DR N-PSDB; ACA58405.
XX
XX New isolated nucleic acid encoding a PRO polypeptide, e.g. PRO245 or
PT PRO1868, useful in molecular biology, chromosome and gene mapping, in
PT generating antisense RNA and DNA, and in gene therapy -
XX
XX Claim 12; Fig 50; 474pp; English.
XX
XX The present invention relates to the isolation of novel human secreted
CC and transmembrane proteins (PRO polypeptides), and the polynucleotide
CC sequences encoding them. The polynucleotide sequences are useful in
CC molecular biology, as hybridisation probes, in chromosome and gene
CC mapping, in generating antisense RNA and DNA, and in gene therapy. The
CC polynucleotide sequences may also be used in preparing PRO polypeptides
CC by recombinant techniques, and in generating either transgenic animals
CC or knock-out animals which, in turn, are useful in the development and
CC screening of therapeutically useful reagents. The PRO polypeptides or
CC their antibodies are useful in preparing a medicament for treating a
CC condition responsive to the polypeptide or antibody, such as cancer,
CC Alzheimer's disease or ischaemia, and in various diagnostic assays.
CC ABU71445-ABU71505 represent human PRO polypeptides of the invention.
XX
XX Sequence 382 AA;
SQ
Query Match 98.7%; Score 1086; DB 24; Length 382;
Best Local Similarity 96.2%; Pred. No. 2.8e-102;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 1 RLLS-----GQVCRGQTQPCVKVIYFHDTSRLNFEEAKEACRRDGGOLVSTSE 52
Db 25 RLLSASDLRLRGQVCRGQTQPCVKVIYFHDTSRLNFEEAKEACRRDGGOLVSTSE 84
QY 53 DEQKLEKFIENLLPSGDFWIGLRRREKQSNSTACQDLIYAWTDGSIQFRNMYVDEPS 112
Db 85 DEQKLEKFIENLLPSGDFWIGLRRREKQSNSTACQDLIYAWTDGSIQFRNMYVDEPS 144
QY 113 CGSEVCVMYHQPAPAGIGGPMFQWNNDRCKMKNFKICKYSDKPAVPSRAEGEETE 172
Db 145 CGSEVCVMYHQPAPAGIGGPMFQWNNDRCKMKNFKICKYSDKPAVPSRAEGEETE 204
QY 173 LTPVLPETQEDAKTKFKESREAAALNLAY 203
Db 205 LTPVLPETQEDAKTKFKESREAAALNLAY 235
RESULT 13
ID ABU71914
XX ABU71914 standard; Protein; 382 AA.
AC ABU71914;

```

```

XX DT 12-JUN-2003 (first entry)
XX DE Human secreted/transmembrane protein PRO234.
XX KW Human; secreted protein; transmembrane protein; PRO;
XX KW gene therapy; chromosome identification; chromosome marker.
XX OS Homo sapiens.
XX FN US2003003530-A1.
XX PD 02-JAN-2003.
XX PF 11-JUL-2001; 2001US-0904011.
XX PR 10-SEP-1998; 98WO-US18824.
XX PR 14-SEP-1998; 98WO-US19177.
XX PR 16-SEP-1998; 98WO-US19330.
XX PR 17-SEP-1998; 98WO-US19437.
XX PR 01-DEC-1998; 98WO-US25108.
XX PR 08-SEP-1999; 99WO-US20594.
XX PR 13-SEP-1999; 99WO-US20944.
XX PR 15-SEP-1999; 99WO-US21090.
XX PR 15-SEP-1999; 99WO-US21547.
XX PR 05-OCT-1999; 99WO-US23089.
XX PR 29-NOV-1999; 99WO-US28214.
XX PR 30-NOV-1999; 99WO-US28313.
XX PR 01-DEC-1999; 99WO-US28301.
XX PR 02-DEC-1999; 99WO-US28564.
XX PR 02-DEC-1999; 99WO-US28565.
XX PR 16-DEC-1999; 99WO-US30095.
XX PR 20-DEC-1999; 99WO-US30911.
XX PR 20-DEC-1999; 99WO-US30999.
XX PR 05-JAN-2000; 2000WO-US00219.
XX PR 11-FEB-2000; 2000WO-US03565.
XX PR 22-FEB-2000; 2000WO-US04414.
XX PR 02-MAR-2000; 2000WO-US05004.
XX PR 02-MAR-2000; 2000WO-US05841.
XX PR 30-MAR-2000; 2000WO-US07377.
XX PR 30-MAR-2000; 2000WO-US08439.
XX PR 22-MAY-2000; 2000WO-US14042.
XX PR 02-JUN-2000; 2000WO-US15264.
XX PR 28-JUL-2000; 2000WO-US20710.
XX PR 24-AUG-2000; 2000WO-US23328.
XX PR 17-SEP-1997; 97US-059113P.
XX PR 17-SEP-1997; 97US-059115P.
XX PR 17-SEP-1997; 97US-059117P.
XX PR 17-SEP-1997; 97US-059119P.
XX PR 17-SEP-1997; 97US-059121P.
XX PR 17-SEP-1997; 97US-059122P.
XX PR 18-SEP-1997; 97US-059184P.
XX PR 18-SEP-1997; 97US-059263P.
XX PR 18-SEP-1997; 97US-059266P.
XX PR 15-OCT-1997; 97US-062125P.
XX PR 17-OCT-1997; 97US-062285P.
XX PR 17-OCT-1997; 97US-062287P.
XX PR 21-OCT-1997; 97US-063486P.
XX PR 24-OCT-1997; 97US-062814P.
XX PR 24-OCT-1997; 97US-062816P.
XX PR 24-OCT-1997; 97US-063045P.
XX PR 24-OCT-1997; 97US-063120P.
XX PR 24-OCT-1997; 97US-063121P.
XX PR 24-OCT-1997; 97US-063127P.
XX PR 24-OCT-1997; 97US-063128P.
XX PR 27-OCT-1997; 97US-063327P.
XX PR 27-OCT-1997; 97US-063329P.
XX PR 28-OCT-1997; 97US-063541P.
XX PR 28-OCT-1997; 97US-063542P.
XX PR 28-OCT-1997; 97US-063544P.
XX PR 28-OCT-1997; 97US-063549P.
XX PR 28-OCT-1997; 97US-063550P.
XX PR 28-OCT-1997; 97US-063564P.

```

PR 29-OCT-1997; 97US-0634335P.
PR 29-OCT-1997; 97US-063704P.
PR 29-OCT-1997; 97US-063732P.
PR 29-OCT-1997; 97US-063733P.
PR 29-OCT-1997; 97US-063734P.
PR 29-OCT-1997; 97US-063735P.
PR 29-OCT-1997; 97US-063738P.
PR 29-OCT-1997; 97US-064215P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 03-NOV-1997; 97US-064248P.
PR 07-NOV-1997; 97US-064809P.
PR 12-NOV-1997; 97US-065186P.
PR 17-NOV-1997; 97US-065846P.
PR 18-NOV-1997; 97US-065693P.
PR 21-NOV-1997; 97US-066120P.
PR 21-NOV-1997; 97US-066364P.
PR 24-NOV-1997; 97US-066453P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066511P.
PR 24-NOV-1997; 97US-066770P.
PR 24-NOV-1997; 97US-066772P.
PR 18-SEP-2000; 2000US-0665350.
XX
XX (GETH) GENENTECH INC.
PA
XX Ashkenazi A, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kljavin LJ;
PI Mather JP, Pan J, Raoni NF, Roy MA, Stewart TA, Tumas D;
PI Williams PM, Wood WI;
XX
XX WPI; 2003-329602/31.
DR N-PSDB; ACA60112.
DR
XX New transmembrane polypeptides and nucleic acids encoding the
PT polypeptides, useful in gene therapy, in chromosome identification, as
PT chromosome markers, in generating probes and in tissue typing
XX
XX Claim 12; Fig 50; 484pp; English.
XX
CC The invention relates to an isolated nucleic acid with at least 80%
CC nucleic acid sequence identity to a nucleotide sequence encoding one of
CC 61 secreted/transmembrane polypeptides, or PRO polypeptides or encoding a
CC PRO protein extracellular domain. Also included are a vector comprising
CC the PRO nucleic acid, a host cell comprising the vector, producing a PRO
CC polypeptide (by culturing the host cell for the expression of the PRO
CC polypeptide, and recovering the PRO polypeptide from the cell culture),
CC an isolated PRO polypeptide (having at least 80% sequence identity
CC to: (a) an amino acid sequence selected from the 61 PRO proteins;
CC (b) an amino acid sequence encoded by a nucleic acid molecule deposited
CC with an ATCC number (detailed in the specification); or (c) an
CC extracellular domain of a PRO polypeptide or to a PRO polypeptide lacking
CC its associated signal peptide), a chimeric molecule comprising a PRO
CC polypeptide of fused to a heterologous amino acid sequence, an anti-PRO
CC antibody, detecting a PRO245 or PRO1868 in a sample suspected of
CC containing the polypeptide, linking a bioactive molecule to a cell
CC expressing a PRO245 or PRO1868 and modulating at least one biological
CC activity of a cell expressing a PRO245 or PRO1868. Nucleic acids which
CC encode PRO can be used to generate either transgenic animals or knock-out
CC animals which may be used in the development and screening of
CC therapeutically useful reagents. The nucleic acids may also be used in
CC gene therapy, in chromosome identification, as chromosome markers, or in
CC generating probes. The PRO polypeptides are useful as molecular markers
CC for protein electrophoresis, and the isolated nucleic acids may be used
CC for recombinantly expressing those markers. The PRO polypeptides and
CC nucleic acids may also be used in tissue typing. Anti-PRO antibodies
CC are useful in diagnostic assays for PRO, and in affinity purification
CC of PRO from recombinant cell culture or natural sources. The
CC present sequence represents a PRO protein.
XX
XX Sequence 382 AA;

Query Match 98.7%; Score 1086; DB 24; Length 382;

Best Local Similarity 96.2%; Pred. No. 2.8e-102;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 1 RLSS-----GQVCRGGTORPCYKVIYFHDTSRLNFEFAKEACRRDGGQLVSESE 52
DB 25 RLSSASDDLRLGQGPVCRGGTORPCYKVIYFHDTSRLNFEFAKEACRRDGGQLVSESE 84
QY 53 DEQKLEKFIENLLPSDGFWGLRRREKQSNSTACQDLYAWTDGSGISQFRNYYVDEPS 112
DB 85 DEQKLEKFIENLLPSDGFWGLRRREKQSNSTACQDLYAWTDGSGISQFRNYYVDEPS 144
QY 113 CGSEVCVMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETE 172
DB 145 CGSEVCVMYHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETE 204
QY 173 LTPVLPEETQEDAKKTFKESREAAALNLAY 203
DB 205 LTPVLPEETQEDAKKTFKESREAAALNLAY 235
RESULT 14
ABU65578
ID ABU65578 standard; Protein; 382 AA.
XX
AC ABU65578;
XX
DT 19-MAY-2003 (first entry)
XX
DE Human secreted/transmembrane protein, SEQ ID 20.
XX
KW Human; PRO; secreted protein; transmembrane protein;
KW cytostatic; antiarthritic; osteopathic; adrenal tumour; lung tumour;
KW colon tumour; breast tumour; prostate tumour; rectal tumour;
KW cervical tumour; liver tumour; TNF-alpha release; arthritis;
KW tumour necrosis factor alpha; chondrocyte cell; bone disorder;
KW cartilage disorder; sports injury.
XX
OS Homo sapiens.
XX
PN US2003036156-A1.
XX
PD 20-FEB-2003.
XX
PF 02-JUL-2002; 2002US-0188767.
XX
PR 16-SEP-1998; 98WO-US19330.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 08-MAR-1999; 99WO-US05028.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28551.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.

PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 28-AUG-2001; 2001WO-US27099.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 29-OCT-1997; 97US-063564P.
PR 31-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069425P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082588P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083455P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.

PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 09-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089658P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 22-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090696P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 02-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 18-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097952P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 01-SEP-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.
PR 02-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.

```
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.

Query Match 98.7%; Score 1086; DB 24; Length 382;
Best Local Similarity 96.2%; Pred. No. 2.8e-102;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 RLLS-----GQVCRGSGTORPCYKVIYFHDTSRLNFEFAKEACRDGGOLVSIESE 52
Db 25 RLLASDLRLGGQPCVCRGTQPCYKVIYFHDTSRLNFEFAKEACRDGGOLVSIESE 84

QY 53 DEQKLIKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPS 112
Db 85 DEQKLIKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPS 144

QY 113 CGSEVCVMYHQPASAPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSRAEGEETE 172
Db 145 CGSEVCVMYHQPASAPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSRAEGEETE 204

QY 173 LTPVLPPEETOEDAKTKPKESREAAINLAY 203
Db 205 LTPVLPPEETOEDAKTKPKESREAAINLAY 235

RESULT 15
ABU65911
ID ABU65911 standard; Protein; 382 AA.
XX AC ABU65911;
XX DT 20-MAY-2003 (first entry)
XX DE Novel human secreted and transmembrane protein PRO234.
XX KW Human; secreted protein; transmembrane protein; cytotstatic;
KW Gene Therapy; TNF-Agonist-Alpha; chondrocyte stimulator; tumour;
KW adrenal tumour; lung tumour; colon tumour; breast tumour;
KW prostate tumour; rectal tumour; cervical tumour; liver tumour.
XX OS Homo sapiens.
XX PN US2003036157-A1.
XX PD 20-FEB-2003.
XX PF 02-JUL-2002; 2002US-0188769.
XX PR 16-SEP-1998; 98WO-US19330.
PR 07-OCT-1998; 98WO-US21141.
PR 01-DEC-1998; 98WO-US25108.
PR 08-MAR-1999; 99WO-US05028.
PR 14-MAY-1999; 99WO-US10733.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 01-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28551.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 28-JUL-2000; 2000WO-US20710.

PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 29-AUG-2001; 2001WO-US27099.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063321P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 29-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069425P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083495P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 15-MAY-1998; 98US-085700P.
PR 18-MAY-1998; 98US-086023P.
PR 22-MAY-1998; 98US-086392P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087098P.
PR 28-MAY-1998; 98US-087208P.
PR 02-JUN-1998; 98US-087609P.
PR 02-JUN-1998; 98US-087759P.
PR 03-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:04:13 ; Search time 29.1819 Seconds
(without alignments)
1120.478 Million cell updates/sec

Title: US-09-887-855-5

Perfect score: 1115

Sequence: 1 ATGRLLSGPVCVGGTQRPC.....EEDAKTKFKSREAAALNLAY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 1115 | 100.0 | 374 | 21 | AA1993948 |
| 2 | 1115 | 100.0 | 374 | 21 | AA1993948 |
| 3 | 1115 | 100.0 | 374 | 23 | ABG66680 |
| 4 | 1115 | 100.0 | 374 | 23 | ABG90203 |
| 5 | 1115 | 100.0 | 387 | 22 | AA1993948 |
| 6 | 1105 | 99.1 | 374 | 21 | AA1993948 |
| 7 | 1101 | 98.7 | 382 | 22 | AA1993948 |
| 8 | 1101 | 98.7 | 382 | 22 | AA1993948 |
| 9 | 1101 | 98.7 | 382 | 22 | AA1993948 |

| | | | | | | |
|----|-------|------|-----|----|-----------|--------------------|
| 10 | 1101 | 98.7 | 382 | 24 | ABU69645 | Novel human secret |
| 11 | 1101 | 98.7 | 382 | 24 | ABU71121 | Human PRO234 prote |
| 12 | 1101 | 98.7 | 382 | 24 | ABU71468 | Human PRO polypept |
| 13 | 1101 | 98.7 | 382 | 24 | ABU71914 | Human secreted/tra |
| 14 | 1101 | 98.7 | 382 | 24 | ABU65578 | Human secreted/tra |
| 15 | 1101 | 98.7 | 382 | 24 | ABU65911 | Novel human secret |
| 16 | 1101 | 98.7 | 382 | 24 | ABU67368 | Human secreted pro |
| 17 | 1101 | 98.7 | 382 | 24 | ABU67415 | Human secreted/tra |
| 18 | 1101 | 98.7 | 382 | 24 | ABU64522 | Human secreted/tra |
| 19 | 1101 | 98.7 | 382 | 24 | ABU65273 | Human PRO polypept |
| 20 | 1101 | 98.7 | 382 | 24 | ABU58409 | Human PRO polypept |
| 21 | 1101 | 98.7 | 382 | 24 | ABU55945 | Human secreted/tra |
| 22 | 1101 | 98.7 | 382 | 24 | ABU56940 | Human PRO polypept |
| 23 | 1101 | 98.7 | 382 | 24 | ABU54370 | Human secreted/tra |
| 24 | 1101 | 98.7 | 382 | 24 | ABU10519 | Human secreted/tra |
| 25 | 561.5 | 50.4 | 273 | 21 | AA18913 | A novel polypeptid |
| 26 | 561.5 | 50.4 | 273 | 22 | AAU12441 | Human PRO1890 poly |
| 27 | 561.5 | 50.4 | 273 | 22 | AA1873309 | Human C-type lecti |
| 28 | 561.5 | 50.4 | 273 | 22 | AA1887609 | Human PRO1890. Ho |
| 29 | 561.5 | 50.4 | 273 | 23 | ABG95934 | Human secreted/tra |
| 30 | 561.5 | 50.4 | 273 | 23 | ABG95554 | Human angiogenesis |
| 31 | 561.5 | 50.4 | 273 | 23 | ABG84948 | Human PRO1890 prot |
| 32 | 561.5 | 50.4 | 273 | 24 | ABU69084 | Human PRO polypept |
| 33 | 561.5 | 50.4 | 273 | 24 | ABU69107 | Human PRO polypept |
| 34 | 561.5 | 50.4 | 273 | 24 | ABU71589 | Human secreted pol |
| 35 | 561.5 | 50.4 | 273 | 24 | ABU72035 | Novel human secret |
| 36 | 561.5 | 50.4 | 273 | 24 | ABU66839 | Human PRO polypept |
| 37 | 561.5 | 50.4 | 273 | 24 | ABU67115 | Human secreted/tra |
| 38 | 561.5 | 50.4 | 273 | 24 | ABU59920 | Novel secreted and |
| 39 | 561.5 | 50.4 | 273 | 24 | ABU69211 | Human polypeptide |
| 40 | 504.5 | 45.2 | 232 | 22 | AA194192 | Human protein sequ |
| 41 | 504.5 | 45.2 | 232 | 22 | AA194192 | Human liver peptid |
| 42 | 275 | 24.7 | 102 | 22 | ABG51278 | Protein #3787 enco |
| 43 | 275 | 24.7 | 102 | 22 | ABG51278 | Peptide #3787 enco |
| 44 | 275 | 24.7 | 102 | 22 | AA1993948 | Human secreted pro |
| 45 | 240 | 21.5 | 81 | 21 | AA1993948 | |

ALIGNMENTS

RES::T 1
AA1993948
ID AA1993948 standard; Protein; 374 AA.
XX AA1993948;
XX 03-OCT-2000 (first entry)
DT Amino acid sequence of a lectin ss3939 polypeptide.
DE Human; lectin ss3939; chromosome 11; gene therapy.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Key 1..21
FT Peptide /note= "signal peptide"
FT Domain 22..227
FT Domain /note= "extracellular coding region"
FT Domain 228..248
FT Domain /note= "predicted transmembrane domain"
FT Domain 249..374
FT Domain /note= "predicted cytoplasmic or intracellular domain"

XX WO2000039296-A1.
XX 06-JUL-2000.
XX Human protein sequ
XX 22-DEC-1999; 99WO-US30523.
XX 23-DEC-1998; 98US-0113820.

| PA | (IMMV) IMMUNEX CORP. | Peptide | 1..24 |
|----------|---|--|----------|
| XX | Anderson DA; | Protein | 25..374 |
| XX | WPI: 2000-452394/39. | /note= "Mature human extracellular matrix and cell | |
| DR | N-PSDB; AAA57302. | adhesion molecule (XMAD)" | |
| XX | ss3939 nucleic acids, polypeptides and antibodies, useful for | Domain | 46..63 |
| PT | identifying human chromosome 11 and diseases associated with it - | Domain | 163..176 |
| XX | Claim 12; Page 8; 73pp; English. | Domain | 224..247 |
| PS | | Domain | 328..348 |
| XX | | /note= "Transmembrane motif" | |
| CC | The present sequence represents a human lectin ss3939 polypeptide. The | /note= "Transmembrane motif" | |
| CC | polynucleotide sequence is a source of probes, which may be used | WO200142285-A2. | |
| CC | to identify nucleic acids encoding ss3939 proteins, to identify human | 14-JUN-2001. | |
| CC | chromosome number 11, to map genes on human chromosome number 11, to | 05-DEC-2000; 2000WO-US32990. | |
| CC | identify diseases associated with chromosome 11, as single-stranded | 10-DEC-1999; 99US-0172852. | |
| CC | sense or antisense oligonucleotides to inhibit expression of | 16-DEC-1999; 99US-0172354. | |
| CC | polypeptides encoded by the ss3939 gene, and for gene therapy. The | (INCY-) INCYTE GENOMICS INC. | |
| CC | ss3939 polypeptides may be useful for developing treatments for | | |
| CC | diseases (none specified) associated with defective or insufficient | | |
| CC | amounts of the polypeptides. The antibodies may be useful for | | |
| CC | detecting the presence of ss3939 polypeptides. | | |
| XX | | | |
| SQ | Sequence 374 AA; | | |
| | Query Match 100.0%; Score 1115; DB 21; Length 374; | | |
| | Best Local Similarity 100.0%; Pred. No. 3.5e-105; | | |
| | Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | |
| QY | 1 ATGRLISGQPVCGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSISEDEKL 60 | | |
| Db | 22 ATGRLISGQPVCGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSISEDEKL 81 | | |
| QY | 61 IEFNIENLPSDGFWIGLRRREKOSNSTACODLYAWTDGSIQFRNMYVDEPSCGSEV 120 | | |
| Db | 82 IEFNIENLPSDGFWIGLRRREKOSNSTACODLYAWTDGSIQFRNMYVDEPSCGSEV 141 | | |
| QY | 121 CVVMYHOPSPAGIGGYPFQWMDRCNNKNNFICKYSDEKPAVPSREAGEBETELTPV 180 | | |
| Db | 142 CVVMYHOPSPAGIGGYPFQWMDRCNNKNNFICKYSDEKPAVPSREAGEBETELTPV 201 | | |
| QY | 181 LPBETQEDAKTKFESREAAALNAY 206 | | |
| Db | 202 LPBETQEDAKTKFESREAAALNAY 227 | | |
| RESULT 2 | | | |
| AAE03651 | | | |
| ID | AAE03651 standard; Protein; 374 AA. | | |
| XX | AC | | |
| XX | AAE03651; | | |
| DT | 06-AUG-2001 (first entry) | | |
| XX | | | |
| DE | Human extracellular matrix and cell adhesion molecule-15 (XMAD-15). | | |
| XX | | | |
| KW | Human; extracellular matrix and cell adhesion molecule; XMAD; | | |
| KW | gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia; | | |
| KW | Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy; | | |
| KW | sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma; | | |
| KW | inflammatory disorder; acquired immune deficiency syndrome; AIDS; | | |
| KW | Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma; | | |
| KW | Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer; | | |
| KW | glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis; | | |
| KW | osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis; | | |
| KW | infection; cell proliferative disorder; actinic keratosis; myeloma; | | |
| KW | arteriosclerosis; neutropenic; anticonvulsant; antithyroid; nephrotropic; | | |
| KW | neuroprotective; dermatological. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| PH | Key Location/Qualifiers | | |
| | Query Match 100.0%; Score 1115; DB 22; Length 374; | | |
| | Best Local Similarity 100.0%; Pred. No. 3.5e-105; | | |
| | Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | |

QY 1 ATGRLLSGQPVCRGGTQPCVKIYFHDTSRRLLNFEEAKACRRDGGQQLVSIIEDEQKL 60
DB 22 ATGRLLSGQPVCRGGTQPCVKIYFHDTSRRLLNFEEAKACRRDGGQQLVSIIEDEQKL 81
QY 61 IEKFIENLLPSDGDGFWIGLRREEKQSNSTACQDLVYAWTDGSIQFRNWWVDEPSCGSEV 120
DB 82 IEKFIENLLPSDGDGFWIGLRREEKQSNSTACQDLVYAWTDGSIQFRNWWVDEPSCGSEV 141
QY 121 CVVMYHQPAPAGIGGYPYFQWMDRCNMKNFICKYSDEKPAVPSREAGEETEELTPV 180
DB 142 CVVMYHQPAPAGIGGYPYFQWMDRCNMKNFICKYSDEKPAVPSREAGEETEELTPV 201
QY 181 LPETQEDAKKTFKESREAAALNLAY 206
DB 202 LPETQEDAKKTFKESREAAALNLAY 227

RESULT 3
ABG66680
ID ABG66680 standard; Protein; 374 AA.
XX AC ABG66680;
XX DT 30-AUG-2002 (first entry)
XX DE Human novel polypeptide #15.
XX KW Human; inflammatory condition; shock; sepsis; immune response;
KW cancer; wound healing; central nervous system disease; haematopoiesis;
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW bone degenerative disorder; periodontal disease; reperfusion injury;
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW allergic condition; thrombolysis; thrombosis; coagulation disorder;
KW fungal infection.
XX OS Homo sapiens.
XX PN WO200244340-A2.
XX PD 06-JUN-2002.
XX PF 30-NOV-2001; 2001WO-US47004.
XX PR 30-NOV-2000; 2000US-0028952.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI Yamazaki V, Ujwal ML, Drmanac RT;
XX DR WPI; 2002-508509/54.
XX DR N-PSDB; ABK94904.
XX PT Novel nucleic acids and polypeptides for diagnosis, treatment of
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT disorders, cancer and promoting wound healing -
XX PS Claim 10; Page 579-580; 672pp; English.
XX CC The invention relates to human novel polynucleotides and associated
CC polypeptides. The polynucleotides and polypeptides are useful for
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC and cancer and for promoting wound healing. The sequences are used to
CC induce the proliferation of neural cells and regeneration of nerve and
CC brain tissue, and are useful for the treatment of central and peripheral
CC nervous system diseases and neuropathies, such as Alzheimer's disease,
CC Parkinson's disease, Huntington's disease and amyotrophic lateral
CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid

CC cell disorders and platelet disorders such as thrombocytopenia,
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC and coagulation disorders. Sequences ABG66666-ABG666758 represent human
CC novel polypeptides of the invention.
XX SQ

Query Match 100.0%; Score 1115; DB 23; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.5e-105;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGRLLSGQPVCRGGTQPCVKIYFHDTSRRLLNFEEAKACRRDGGQQLVSIIEDEQKL 60
DB 22 ATGRLLSGQPVCRGGTQPCVKIYFHDTSRRLLNFEEAKACRRDGGQQLVSIIEDEQKL 81
QY 61 IEKFIENLLPSDGDGFWIGLRREEKQSNSTACQDLVYAWTDGSIQFRNWWVDEPSCGSEV 120
DB 82 IEKFIENLLPSDGDGFWIGLRREEKQSNSTACQDLVYAWTDGSIQFRNWWVDEPSCGSEV 141
QY 121 CVVMYHQPAPAGIGGYPYFQWMDRCNMKNFICKYSDEKPAVPSREAGEETEELTPV 180
DB 142 CVVMYHQPAPAGIGGYPYFQWMDRCNMKNFICKYSDEKPAVPSREAGEETEELTPV 201
QY 181 LPETQEDAKKTFKESREAAALNLAY 206
DB 202 LPETQEDAKKTFKESREAAALNLAY 227
Human polypeptide SEQ ID NO 2579.

RESULT 4
ABB90203
ID ABB90203 standard; Protein; 374 AA.
XX AC ABB90203;
XX DT 24-MAY-2002 (first entry)
XX DE Human polypeptide SEQ ID NO 2579.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX OS Homo sapiens.
XX PN WO200190304-A2.
XX PD 29-NOV-2001.
XX PF 18-MAY-2001; 2001WO-US16450.
XX PR 19-MAY-2000; 2000US-205515P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PI Birse CE, Rosen CA;
XX WPI; 2002-122018/16.
XX DR N-PSDB; ABL90612.
XX PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -

```

XX PS Claim 11; SEQ ID NO 2579; 2081pp + Sequence Listing; English.
XX CC
XX CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
XX CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
XX CC medical conditions e.g. by protein or gene therapy. The genes are
XX CC isolated from a range of human tissues disclosed in the specification.
XX CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX CC and parasitic infections.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ
XX SQ Sequence 374 AA;
XX Query Match 100.0%; Score 1115; DB 23; Length 374;
XX Best Local Similarity 100.0%; Pred. No. 3.5e-105;
XX Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 ATGRLSGQPVCRGGTQPCVKYVFHDTSRRLNFEAKEACRRDGGQGVISEDEOKL 60
XX DB 22 ATGRLSGQPVCRGGTQPCVKYVFHDTSRRLNFEAKEACRRDGGQGVISEDEOKL 81
XX
XX QY 61 IEKFTIENLLPSDGFWMGLRRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 120
XX DB 82 IEKFTIENLLPSDGFWMGLRRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 141
XX
XX QY 121 CVVMYHQPSPAGIGGYPYMFQWDDRCNMKNFICKYSDKPAVPSRAEGEETLTPV 180
XX DB 142 CVVMYHQPSPAGIGGYPYMFQWDDRCNMKNFICKYSDKPAVPSRAEGEETLTPV 201
XX
XX QY 181 LPEETQEDAKKTFKESREAAALNLAY 206
XX DB 202 LPEETQEDAKKTFKESREAAALNLAY 227
XX
XX RESULT 5
XX AAM25796
XX ID AAM25796 standard; Protein; 387 AA.
XX AC AAM25796;
XX
XX XX 16-OCT-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:1311.
XX
XX KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
XX KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
XX KW antibacterial; endocrine; cardiant; central nervous system; virucide;
XX KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
XX KW antiaggregant; haemostatic; vulnerary; antitumor; osteopathic; eczema;
XX KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
XX KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
XX KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
XX KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
XX KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
XX KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
XX KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
XX KW allergic rhinitis; diabetes; multiple sclerosis; depression;
XX KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
XX KW neurological disorder.
XX OS Homo sapiens.
XX XX

```

```

PN WO200153455-A2.
XX PD 26-JUL-2001.
XX PF 22-DEC-2000; 2000WO-US35017.
XX PR 23-DEC-1999; 99US-0471275.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-457603/49.
XX DR N-PSDB; AAH99737.
XX PT Isolated human polynucleotides encoding polypeptides, useful for the
XX PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX PS Claim 20; Page 272; 1217pp; English.
XX
XX CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XX CC AAM25963. The proteins can have activities based on the tissues and
XX CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
XX CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
XX CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
XX CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;
XX CC antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;
XX CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
XX CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX CC encoding them can be used in gene therapy, antisense therapy and vaccine
XX CC production. The proteins and polynucleotides are useful for screening for
XX CC agonists or antagonists of a protein and for the treatment and diagnosis
XX CC of disorders associated with the activity of a protein e.g. inflammation,
XX CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XX CC osteoporosis, severe combined immunodeficiency, eczema, allergic
XX CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX CC neurological disorders.
XX SQ
XX SQ Sequence 387 AA;
XX Query Match 100.0%; Score 1115; DB 22; Length 387;
XX Best Local Similarity 100.0%; Pred. No. 3.6e-105;
XX Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 ATGRLSGQPVCRGGTQPCVKYVFHDTSRRLNFEAKEACRRDGGQGVISEDEOKL 60
XX DB 35 ATGRLSGQPVCRGGTQPCVKYVFHDTSRRLNFEAKEACRRDGGQGVISEDEOKL 94
XX
XX QY 61 IEKFTIENLLPSDGFWMGLRRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 120
XX DB 95 IEKFTIENLLPSDGFWMGLRRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 154
XX
XX QY 121 CVVMYHQPSPAGIGGYPYMFQWDDRCNMKNFICKYSDKPAVPSRAEGEETLTPV 180
XX DB 155 CVVMYHQPSPAGIGGYPYMFQWDDRCNMKNFICKYSDKPAVPSRAEGEETLTPV 214
XX
XX QY 181 LPEETQEDAKKTFKESREAAALNLAY 206
XX DB 215 LPEETQEDAKKTFKESREAAALNLAY 240
XX
XX RESULT 6
XX AAY91490
XX ID AAY91490 standard; Protein; 374 AA.
XX AC AAY91490;
XX XX

```

29-JUN-2000 (first entry)
Human secreted protein sequence encoded by gene 40 SEQ ID NO:163.
Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
antiHIV; antiinflammatory; neutrotic; neuroprotective; antiAllergic;
osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
antipsoriatic; cardiac; gene therapy; cancer; neurological disorder;
immune disease; inflammation; blood disorder; tumour.
Homo sapiens.
WO200006698-A1.
10-FEB-2000.
29-JUL-1999; 99WO-US17130.
30-JUL-1998; 98US-0094657.
05-AUG-1998; 98US-0095486.
06-AUG-1998; 98US-0095454.
06-AUG-1998; 98US-0095455.
12-AUG-1998; 98US-0096319.
(HUMA-) HUMAN GENOME SCI INC.
Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
WPI; 2000-195282/17.
N-PSDB; AAA26385.
New isolated human genes and the secreted polypeptides they encode,
useful for diagnosis and treatment of e.g. cancers, neurological
disorders, immune diseases, inflammation or blood disorders
Claim 11; Page 483-484; 634pp; English.
The polynucleotide sequences given in AAA26346 to AAA26458 encode the
human secreted proteins given in AAY91451 to AAY91691. The human secreted
proteins can have activities based on the tissues and cells they are
expressed in. Examples of the activities are: cytostatic;
immunosuppressive; antiHIV; antiinflammatory; cytostatic;
antiAllergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
antiasthma; antipsoriatic; and cardiac. The polynucleotides and their
corresponding secreted proteins are useful for preventing, treating or
ameliorating medical conditions, e.g. by protein or gene therapy. Also
pathological conditions can be diagnosed by determining the amount of the
proteins in a sample or by determining the presence of mutations in the
polynucleotides. Specific uses are described for each of the
polynucleotides, based on which tissues they are most highly expressed
in, and include developing products for the diagnosis or treatment of
cancer, tumours, neurodegenerative disorders, developmental abnormalities
and foetal deficiencies, blood disorders, diseases of the immune system,
autoimmune diseases, hepatic and renal disease, inflammation,
allergies, Alzheimer's and behavioural disorders, schizophrenia,
osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
cardiovascular disorders, reproductive disorders, gastrointestinal
disorders, respiratory disorders and metabolic disorders. The proteins
or polynucleotides can also be used as food additives or preservatives.
The proteins are also useful for identifying their binding partners.
AAA26337 to AAA26345 and AAY91450 are sequences used in the
exemplification of the present invention.

Query Match 99.1%; Score 1105; DB 21; Length 374;
Best Local Similarity 99.0%; Pred. No. 3.7e-104;
Matches 204; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGRLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEAKCRDGGQLVSEDEQKL 60

Db 22 ATGRLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEAKCRDGGQLVSEDEQKL 81
QY 61 IEKFENLLPSDGDGFWIGLRRREKOSNSTACODLYAWTDGSIQFRNWWVDEPSCGSEV 120
Db 82 IEKFENLLPSDGDGFWIGLRRREKOSNSTYCODLYAWTDGSIQFRNWWVDEPSCGSEV 141
QY 121 CVVMYHQPSAPAGIGGPFQWNNDRCNMKNFICKYSDKPAVPSREAGEETEELTPV 180
Db 142 CVVMYHQPSAPAGIGGPFQWNNDRCNMKNFICKYSDKPAVPSREAGEETEELTPV 201
QY 181 LPEETOEDAKKTFKESREAAALNLAY 206
Db 202 LPEETOEDAKKTFKESREAAALNLAY 227
RESULT 7
AAY13367
ID AAY13367 standard; Protein; 382 AA.
AC AAY13367;
DX 25-JUN-1999 (first entry)
XX Amino acid sequence of protein PRO234.
XX Secreted protein; transmembrane protein; human; enterocolitis;
KW Zollinger-Ellison syndrome; gastrointestinal ulceration;
KW congenital microvillus atrophy; skin disease; cell growth;
KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KW Parkinson's disease; Alzheimer's disease; AIDS; neuropathy;
KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
KW anti-thrombotic; wound healing; tissue repair.
OS Homo sapiens.
XX WO9914328-A2.
XX 25-MAR-1999.
XX 16-SEP-1998; 98WO-US19330.
XX 25-NOV-1997; 97US-0066840.
PR 17-SEP-1997; 97US-0059113.
PR 17-SEP-1997; 97US-0059115.
PR 17-SEP-1997; 97US-0059117.
PR 17-SEP-1997; 97US-0059119.
PR 17-SEP-1997; 97US-0059121.
PR 17-SEP-1997; 97US-0059122.
PR 17-SEP-1997; 97US-0059184.
PR 18-SEP-1997; 97US-0059263.
PR 18-SEP-1997; 97US-0059266.
PR 15-OCT-1997; 97US-0062125.
PR 17-OCT-1997; 97US-0062285.
PR 21-OCT-1997; 97US-0063486.
PR 24-OCT-1997; 97US-0062814.
PR 24-OCT-1997; 97US-0062815.
PR 24-OCT-1997; 97US-0063045.
PR 24-OCT-1997; 97US-0063120.
PR 24-OCT-1997; 97US-0063121.
PR 24-OCT-1997; 97US-0063127.
PR 24-OCT-1997; 97US-0063128.
PR 27-OCT-1997; 97US-0063329.
PR 27-OCT-1997; 97US-0063327.
PR 28-OCT-1997; 97US-0063541.
PR 28-OCT-1997; 97US-0063542.
PR 28-OCT-1997; 97US-0063544.
PR 28-OCT-1997; 97US-0063549.
PR 28-OCT-1997; 97US-0063550.
PR 28-OCT-1997; 97US-0063564.
PR 29-OCT-1997; 97US-0063435.
PR 29-OCT-1997; 97US-0063704.

```

PR 29-OCT-1997; 97US-0063732.
PR 29-OCT-1997; 97US-0063738.
PR 29-OCT-1997; 97US-0063734.
PR 29-OCT-1997; 97US-0064215.
PR 29-OCT-1997; 97US-0063735.
PR 31-OCT-1997; 97US-0063870.
PR 31-OCT-1997; 97US-0064103.
PR 03-NOV-1997; 97US-0064248.
PR 07-NOV-1997; 97US-0064809.
PR 12-NOV-1997; 97US-0065186.
PR 17-NOV-1997; 97US-0065846.
PR 18-NOV-1997; 97US-0065693.
PR 21-NOV-1997; 97US-0066120.
PR 21-NOV-1997; 97US-0066364.
PR 24-NOV-1997; 97US-0066772.
PR 24-NOV-1997; 97US-0066466.
PR 24-NOV-1997; 97US-0066770.
PR 24-NOV-1997; 97US-0066511.
PR 24-NOV-1997; 97US-0066453.
XX
PA (GETH ) GENENTECH INC.
XX
XX Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
XX
XX WPI: 1999-229533/19.
XX N-PSDB; AAX52238.
XX
XX New isolated human genes and polypeptides used in, e.g. treatment of
XX gastrointestinal ulceration
XX
XX Claim 12; Fig 50; 320pp; English.
XX
XX AAY13344-403 represent secreted and transmembrane human proteins.
XX The cDNA sequences are obtained from cDNA libraries, prepared from
XX fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
XX The encoded polypeptides have specific uses based on their homology to
XX known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
XX associated with the preservation and maintenance of gastrointestinal
XX mucosa and the repair of acute and chronic mucosal lesions
XX (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
XX ulceration and congenital microvillus atrophy), skin diseases associated
XX with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
XX cancers such as lung squamous cell carcinoma of the vulva and gliomas),
XX potent effects on cell growth and development, diseases related to
XX growth or survival of nerve cells including Parkinson's disease,
XX Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
XX for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
XX as a target for anti-tumor drugs. PRO533 may be used in the treatment
XX of Usher Syndrome or Atrophia areata; PRO269 can be used as an
XX anti-thrombotic agent; PRO287 polypeptides and portions may have
XX therapeutic applications in wound healing and tissue repair; PRO317 can
XX be used for treating problems of the kidney, uterus, endometrium, blood
XX vessels, or related tissue, e.g. in the heart of genital tract.
XX
XX SQ Sequence 382 AA;
XX
XX Query Match 98.7%; Score 1101; DB 20; Length 382;
XX Best Local Similarity 96.3%; Pred. No. 9.6e-104;
XX Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
XX
XX 1 ATGRLLS-----GQVRCRGGTQRCYKVIYFHDTSRLRNFEAEKACRRDGGQLVSI 52
XX |||||
XX 22 ATGRLLSASDLDRGGQPVRCGTQRCYKVIYFHDTSRLRNFEAEKACRRDGGQLVSI 81
XX |||||
XX 53 ESEDEQKLIKFIENLPSDGFWMIGLRREKQSNSTACQDLYAWTDGSIQFRNMYVD 112
XX |||||
XX 82 ESEDEQKLIKFIENLPSDGFWMIGLRREKQSNSTACQDLYAWTDGSIQFRNMYVD 141
XX |||||
XX 113 EPCSGSEVCVMYHQPAPAGIGPYMFQNDRCNMKNFKICKYSDEKPAVPSREAGE 172
XX |||||
XX 142 EPCSGSEVCVMYHQPAPAGIGPYMFQNDRCNMKNFKICKYSDEKPAVPSREAGE 201
XX |||||
XX 173 ETELTPVLPEETQEDAKKTFKESREAAALNLAY 206
XX |||||
XX 202 ETELTPVLPEETQEDAKKTFKESREAAALNLAY 235
XX |||||
XX
XX RESULT 8
XX AAU29033
XX ID AAU29033 standard; Protein; 382 AA.
XX AC AAU29033;
XX XX 18-DEC-2001 (first entry)
XX DE Human PRO polypeptide sequence #10.
XX KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
XX dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
XX blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
XX adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX OS Homo sapiens.
XX PN WO200168848-A2.
XX PD 20-SEP-2001.
XX PF 28-FEB-2001; 2001WO-US06520.
XX
XX 01-MAR-2000; 2000WO-US05601.
XX 02-MAR-2000; 2000WO-US05841.
XX 03-MAR-2000; 2000US-187202P.
XX 06-MAR-2000; 2000US-186968P.
XX 14-MAR-2000; 2000US-189320P.
XX 14-MAR-2000; 2000US-189328P.
XX 15-MAR-2000; 2000WO-US06884.
XX 21-MAR-2000; 2000US-190828P.
XX 21-MAR-2000; 2000US-191007P.
XX 21-MAR-2000; 2000US-191048P.
XX 21-MAR-2000; 2000US-191314P.
XX 28-MAR-2000; 2000US-192655P.
XX 29-MAR-2000; 2000US-193032P.
XX 29-MAR-2000; 2000US-193053P.
XX 30-MAR-2000; 2000WO-US08439.
XX 04-APR-2000; 2000US-194449P.
XX 04-APR-2000; 2000US-194647P.
XX 11-APR-2000; 2000US-195975P.
XX 11-APR-2000; 2000US-196000P.
XX 11-APR-2000; 2000US-196187P.
XX 11-APR-2000; 2000US-196690P.
XX 11-APR-2000; 2000US-196820P.
XX 18-APR-2000; 2000US-198121P.
XX 18-APR-2000; 2000US-198585P.
XX 25-APR-2000; 2000US-199550P.
XX 25-APR-2000; 2000US-199654P.
XX 03-MAY-2000; 2000US-201516P.
XX 17-MAY-2000; 2000WO-US13705.
XX 22-MAY-2000; 2000WO-US14042.
XX 30-MAY-2000; 2000WO-US14941.
XX 02-JUN-2000; 2000WO-US15264.
XX 05-JUN-2000; 2000US-209832P.
XX 22-JUL-2000; 2000WO-US20710.
XX 22-AUG-2000; 2000US-0644848.
XX 24-AUG-2000; 2000WO-US23328.
XX 08-NOV-2000; 2000WO-US30952.
XX 01-DEC-2000; 2000WO-US32678.
XX 20-DEC-2000; 2000WO-US34956.
XX
XX (GETH ) GENENTECH INC.
XX
XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
XX Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX WPI: 2001-602746/68.

```

DR N-PSDB; AAS45934.
 XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -
 XX
 XX
 PS Claim 11; Fig 20; 774pp; English.
 XX
 CC Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
 CC The PRO polypeptides and their associated nucleic acids can be used to
 CC detect the presence of a tumour in a mammal by comparing the level of
 CC expression of a PRO polypeptide in a test sample of cells from the animal
 CC and a control sample of normal cells, whereby a higher level of
 CC expression in the test sample indicates the presence of a tumour in the
 CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
 CC and rabbits but are preferably human. The polypeptides can be used to
 CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
 CC when contacted with it. A specific polypeptide can be used to stimulate
 CC the proliferation or differentiation of chondrocyte cells. The PRO
 CC proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.
 XX
 SQ Sequence 382 AA;

Query Match 98.7%; Score 1101; DB 22; Length 382;
 Best Local Similarity 96.3%; Pred. No. 9.6e-104;
 Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 ATGRLS-----GQVCRGGTORPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSI 52
 DB 22 ATGRLSASDLRLGGQPCVCRGGTORPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSI 81
 QY 53 ESEDEQKLIKFIENLLPSDGFWIGLRRRREKOSNSTACODLYAWTDGTSIQFRNMYVD 112
 DB 82 ESEDEQKLIKFIENLLPSDGFWIGLRRRREKOSNSTACODLYAWTDGTSIQFRNMYVD 141
 QY 113 EPSCGSEVVMYHQPAPAGIGGPFYFQWDDRCNKNFNICKYDEKPAVPSRAEAGE 172
 DB 142 EPSCGSEVVMYHQPAPAGIGGPFYFQWDDRCNKNFNICKYDEKPAVPSRAEAGE 201
 QY 173 ETELTTPVLPEETOEDAKKTFKESREAAALNLAY 206
 DB 202 ETELTTPVLPEETOEDAKKTFKESREAAALNLAY 235

RESULT 9
 AAB80235
 ID AAB80235 standard; Protein; 382 AA.
 XX
 AC AAB80235;
 XX
 XX 24-APR-2001 (first entry)
 XX
 DE Human PRO234 protein.
 XX
 KW Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
 KW antiparkinsonian nootropic; neuroprotective; vulnery; cardiant;
 KW antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;
 KW antiarthritic; antinfertility; antidiabetic; antiviral; diabetes;
 KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
 KW ischaemia; inflammation.
 XX
 OS Homo sapiens.
 XX
 XX WO200104311-A1.
 PN
 XX 18-JAN-2001.
 PD
 XX 22-FEB-2000; 2000WO-US04414.
 PF
 XX
 XX

PR 07-JUL-1999; 99US-0143048.
 PR 26-JUL-1999; 99US-0145698.
 PR 28-JUL-1999; 99US-0146222.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 05-OCT-1999; 99WO-US23089.
 PR 29-NOV-1999; 99WO-US28214.
 PR 30-NOV-1999; 99WO-US28313.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 05-JAN-2000; 99WO-US00219.
 XX
 XX (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
 PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin IJ;
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
 PI Williams PM, Wood WI;
 XX
 DR WPI; 2001-081051/09.
 DR N-PSDB; AAF72396.
 XX

Sixty one nucleic acids encoding PRO polypeptides which are useful in
 the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
 squamous cell carcinoma) and neurodegenerative diseases (e.g.
 Alzheimer's disease) -

Claim 1; Fig 50; 393pp; English.

The present sequence is one of sixty one novel secreted and
 transmembrane PRO polypeptides. The PRO polypeptides are
 useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
 squamous cell carcinoma), gastrointestinal disorders (e.g.
 enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
 Parkinson's disease), wound repair, cardiovascular disorders (e.g.
 endometrial bleeding, angiogenesis, ischaemias such as coronary
 ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
 rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
 diabetes and retinal disorders such as retinitis pigmentosum.
 CC The PRO nucleic acids have applications in molecular biology, including
 CC use as hybridization probes, and in chromosome and gene mapping.
 XX
 SQ Sequence 382 AA;

Query Match 98.7%; Score 1101; DB 22; Length 382;
 Best Local Similarity 96.3%; Pred. No. 9.6e-104;
 Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 ATGRLS-----GQVCRGGTORPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSI 52
 DB 22 ATGRLSASDLRLGGQPCVCRGGTORPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSI 81
 QY 53 ESEDEQKLIKFIENLLPSDGFWIGLRRRREKOSNSTACODLYAWTDGTSIQFRNMYVD 112
 DB 82 ESEDEQKLIKFIENLLPSDGFWIGLRRRREKOSNSTACODLYAWTDGTSIQFRNMYVD 141
 QY 113 EPSCGSEVVMYHQPAPAGIGGPFYFQWDDRCNKNFNICKYDEKPAVPSRAEAGE 172
 DB 142 EPSCGSEVVMYHQPAPAGIGGPFYFQWDDRCNKNFNICKYDEKPAVPSRAEAGE 201
 QY 173 ETELTTPVLPEETOEDAKKTFKESREAAALNLAY 206
 DB 202 ETELTTPVLPEETOEDAKKTFKESREAAALNLAY 235

RESULT 10
 ABU69645
 ID ABU69645 standard; Protein; 382 AA.
 XX

AC ABU69645;
 XX 05-JUN-2003 (first entry)
 XX
 XX Novel human secreted and transmembrane protein PRO234.
 XX
 KW Human; secreted and transmembrane protein; gene therapy; psoriasis;
 KW enterocolitis; gastrointestinal ulceration; skin disease;
 KW keratinocyte differentiation; epithelial cancer; Alzheimer's disease;
 KW squamous cell carcinoma; Parkinson's disease; inflammatory disease;
 KW amyotrophic lateral sclerosis; rheumatoid arthritis; asthma;
 KW multiple sclerosis; organ failure; atherosclerosis; cardiac injury;
 KW infertility; birth defect; premature aging; AIDS; cancer;
 KW diabetic complication; wound repair; tissue re-growth.
 XX
 OS Homo sapiens.
 XX
 XX US2003017463-A1.
 PW
 XX
 XX 23-JAN-2003.
 PD
 XX
 XX 11-JUL-2001; 2001US-0903640.
 PF
 XX
 XX 10-SEP-1998; 98WO-US18824.
 PR 14-SEP-1998; 98WO-US19177.
 PR 16-SEP-1998; 98WO-US19330.
 PR 17-SEP-1998; 98WO-US19437.
 PR 01-DEC-1998; 98WO-US25108.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 05-OCT-1999; 99WO-US23089.
 PR 29-NOV-1999; 99WO-US28214.
 PR 30-NOV-1999; 99WO-US28313.
 PR 01-DEC-1999; 99WO-US28301.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 20-MAR-2000; 2000WO-US07177.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 17-SEP-1997; 97US-059113P.
 PR 17-SEP-1997; 97US-059115P.
 PR 17-SEP-1997; 97US-059117P.
 PR 17-SEP-1997; 97US-059119P.
 PR 17-SEP-1997; 97US-059121P.
 PR 17-SEP-1997; 97US-059122P.
 PR 17-SEP-1997; 97US-059184P.
 PR 18-SEP-1997; 97US-059263P.
 PR 18-SEP-1997; 97US-059266P.
 PR 15-OCT-1997; 97US-062125P.
 PR 17-OCT-1997; 97US-062285P.
 PR 17-OCT-1997; 97US-062287P.
 PR 21-OCT-1997; 97US-063486P.
 PR 24-OCT-1997; 97US-062814P.
 PR 24-OCT-1997; 97US-062816P.
 PR 24-OCT-1997; 97US-063045P.
 PR 24-OCT-1997; 97US-063120P.
 PR 24-OCT-1997; 97US-063121P.
 PR 24-OCT-1997; 97US-063127P.
 PR 24-OCT-1997; 97US-063128P.
 PR 27-OCT-1997; 97US-063327P.
 PR 27-OCT-1997; 97US-063329P.
 PR 28-OCT-1997; 97US-063541P.
 PR 28-OCT-1997; 97US-063542P.
 PR 28-OCT-1997; 97US-063544P.
 PR 28-OCT-1997; 97US-063549P.
 PR 28-OCT-1997; 97US-063550P.
 PR 28-OCT-1997; 97US-063564P.
 PR 29-OCT-1997; 97US-063435P.
 PR 29-OCT-1997; 97US-063704P.
 PR 29-OCT-1997; 97US-063732P.
 PR 29-OCT-1997; 97US-063734P.
 PR 29-OCT-1997; 97US-063735P.
 PR 29-OCT-1997; 97US-063738P.
 PR 29-OCT-1997; 97US-064215P.
 PR 31-OCT-1997; 97US-063870P.
 PR 31-OCT-1997; 97US-064103P.
 PR 03-NOV-1997; 97US-064248P.
 PR 07-NOV-1997; 97US-064809P.
 PR 12-NOV-1997; 97US-065186P.
 PR 17-NOV-1997; 97US-065846P.
 PR 18-NOV-1997; 97US-065693P.
 PR 21-NOV-1997; 97US-066120P.
 PR 21-NOV-1997; 97US-066364P.
 PR 24-NOV-1997; 97US-066453P.
 PR 24-NOV-1997; 97US-066466P.
 PR 24-NOV-1997; 97US-066511P.
 PR 24-NOV-1997; 97US-066770P.
 PR 24-NOV-1997; 97US-066772P.
 PR 25-NOV-1997; 97US-066840P.
 PR 12-DEC-1997; 97US-069425P.
 PR 04-JUN-1998; 98US-088026P.
 PR 10-SEP-1998; 98US-099803P.
 PR 14-SEP-1998; 98US-100262P.
 PR 17-SEP-1998; 98US-100858P.
 PR 13-OCT-1998; 98US-104080P.
 PR 20-NOV-1998; 98US-109304P.
 PR 22-DEC-1998; 98US-113296P.
 PR 07-JUL-1999; 99US-143048P.
 PR 26-JUL-1999; 99US-145698P.
 PR 28-JUL-1999; 99US-146222P.
 PR 18-SEP-2000; 2000US-0665350.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi A, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
 PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kijavlin IJ;
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
 PI Williams PM, Wood WI;
 XX
 DR WPI; 2003-341586/32.
 DR N-PSDB; ACA54920.
 XX
 PT New PRO polypeptides and nucleic acid molecules, useful in diagnosing
 PT or treating inflammatory diseases, organ failure, atherosclerosis,
 PT cardiac injury, infertility, cancer, AIDS, Alzheimer's disease or
 PT Parkinson's disease -
 XX
 PS Claim 12; Fig 50; 473pp; English.
 XX
 CC The invention describes sixty one nucleic acids encoding PRO polypeptides
 CC (secreted and transmembrane). The PRO polypeptides and nucleic acids are
 CC useful in diagnosing or treating enterocolitis, gastrointestinal
 CC ulceration, skin diseases associated with abnormal keratinocyte
 CC differentiation, e.g. psoriasis or epithelial cancers such as squamous
 CC cell carcinoma, Alzheimer's disease, Parkinson's disease, amyotrophic
 CC lateral sclerosis, inflammatory diseases, e.g. rheumatoid arthritis,
 CC asthma or multiple sclerosis, organ failure, atherosclerosis, cardiac
 CC injury, infertility, birth defects, premature aging, AIDS, cancer,
 CC diabetic complications, or mutations in general. The polypeptides are
 CC also useful for wound repair and associated therapies concerned with
 CC re-growth of tissue. The PRO polypeptides and nucleic acid molecules
 CC are also useful in gene therapy, and as molecular weight markers for


```

PR 22-FEB-2000; 2000WO-US04414.
PR 28-JUL-2000; 2000WO-US20710.
PR 24-AUG-2000; 2000WO-US23328.
PR 17-SEP-1997; 97US-059113P.
PR 17-SEP-1997; 97US-059115P.
PR 17-SEP-1997; 97US-059117P.
PR 18-SEP-1997; 97US-059266P.
PR 15-OCT-1997; 97US-062125P.
PR 17-OCT-1997; 97US-062285P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-062814P.
PR 24-OCT-1997; 97US-062816P.
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi A, Botstein D, Desnovers L, Eaton DL, Ferrara N;
XX Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
XX Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kljavin IJ;
XX Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
XX Williams PM, Wood WI;
XX WPI; 2003-361832/34.
XX N-PSDB; ACA58405.
XX
XX New isolated nucleic acid encoding a PRO polypeptide, e.g. PRO245 or
XX PRO1868, useful in molecular biology, chromosome and gene mapping, in
XX generating antisense RNA and DNA, and in gene therapy
XX
XX Claim 12; Fig 50; 474pp; English.
XX
XX The present invention relates to the isolation of novel human secreted
XX and transmembrane proteins (PRO polypeptides), and the polynucleotide
XX sequences encoding them. The polynucleotide sequences are useful in
XX molecular biology, as hybridisation probes, in chromosome and gene
XX mapping, in generating antisense RNA and DNA, and in gene therapy. The
XX polynucleotide sequences may also be used in preparing PRO polypeptides
XX by recombinant techniques, and in generating either transgenic animals
XX or knock-out animals which, in turn, are useful in the development and
XX screening of therapeutically useful reagents. The PRO polypeptides or
XX their antibodies are useful in preparing a medicament for treating a
XX condition responsive to the polypeptide or antibody, such as cancer,
XX Alzheimer's disease or ischaemia, and in various diagnostic assays.
XX ABU71445-ABU71505 represent human PRO polypeptides of the invention.
XX
SQ Sequence 382 AA;
Query Match 98.7%; Score 1101; DB 24; Length 382;
Best Local Similarity 96.3%; Pred. No. 9,6e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
Qy 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSI 81
Qy 53 ESEDEQKLEKFIENLLPSDGFVIGLRRREKOSNSTACODLYAWTDGSIQPRNMYVD 112
Db 82 ESEDEQKLEKFIENLLPSDGFVIGLRRREKOSNSTACODLYAWTDGSIQPRNMYVD 141
Qy 113 EPSCGSEVVCVMYHQPSAPAGIGGYPYMFQWDDRCNNKNNFICKYSDEKPAVPSREAGE 172
Db 142 EPSCGSEVVCVMYHQPSAPAGIGGYPYMFQWDDRCNNKNNFICKYSDEKPAVPSREAGE 201
Qy 173 ETELTTPVLPEETOEDAKTKTFKESREAAANLAY 206
Db 202 ETELTTPVLPEETOEDAKTKTFKESREAAANLAY 235
XX
RESULT 13
ABU71914
ID ABU71914 standard; Protein; 382 AA.
XX
AC ABU71914;

```

```

XX 12-JUN-2003 (first entry)
XX
XX Human secreted/transmembrane protein PRO234.
XX
XX Human; secreted protein; transmembrane protein; PRO;
XX gene therapy; chromosome identification; chromosome marker.
XX
XX Homo sapiens.
XX
XX US2003003530-A1.
XX
XX 02-JAN-2003.
XX
XX 11-JUL-2001; 2001US-0904011.
XX
XX 10-SEP-1998; 98WO-US18824.
XX 14-SEP-1998; 98WO-US19177.
XX 16-SEP-1998; 98WO-US19330.
XX 17-SEP-1998; 98WO-US19437.
XX 01-DEC-1998; 98WO-US25108.
XX 08-SEP-1999; 99WO-US20594.
XX 13-SEP-1999; 99WO-US20944.
XX 15-SEP-1999; 99WO-US21090.
XX 15-SEP-1999; 99WO-US21547.
XX 05-OCT-1999; 99WO-US23089.
XX 29-NOV-1999; 99WO-US28214.
XX 30-NOV-1999; 99WO-US28313.
XX 01-DEC-1999; 99WO-US28301.
XX 02-DEC-1999; 99WO-US28584.
XX 02-DEC-1999; 99WO-US28585.
XX 16-DEC-1999; 99WO-US30095.
XX 20-DEC-1999; 99WO-US30911.
XX 20-DEC-1999; 99WO-US30999.
XX 05-JAN-2000; 2000WO-US00219.
XX 11-FEB-2000; 2000WO-US03565.
XX 22-FEB-2000; 2000WO-US04414.
XX 24-FEB-2000; 2000WO-US05004.
XX 02-MAR-2000; 2000WO-US05841.
XX 20-MAR-2000; 2000WO-US07377.
XX 30-MAR-2000; 2000WO-US08439.
XX 22-MAY-2000; 2000WO-US14042.
XX 02-JUN-2000; 2000WO-US15264.
XX 28-JUL-2000; 2000WO-US20710.
XX 24-AUG-2000; 2000WO-US23328.
XX 17-SEP-1997; 97US-059113P.
XX 17-SEP-1997; 97US-059115P.
XX 17-SEP-1997; 97US-059117P.
XX 17-SEP-1997; 97US-059121P.
XX 17-SEP-1997; 97US-059122P.
XX 17-SEP-1997; 97US-059184P.
XX 18-SEP-1997; 97US-059263P.
XX 18-SEP-1997; 97US-059266P.
XX 15-OCT-1997; 97US-062125P.
XX 17-OCT-1997; 97US-062285P.
XX 17-OCT-1997; 97US-062287P.
XX 21-OCT-1997; 97US-063486P.
XX 24-OCT-1997; 97US-062814P.
XX 24-OCT-1997; 97US-062816P.
XX 24-OCT-1997; 97US-063045P.
XX 24-OCT-1997; 97US-063120P.
XX 24-OCT-1997; 97US-063121P.
XX 24-OCT-1997; 97US-063127P.
XX 24-OCT-1997; 97US-063128P.
XX 27-OCT-1997; 97US-063327P.
XX 27-OCT-1997; 97US-063329P.
XX 28-OCT-1997; 97US-063541P.
XX 28-OCT-1997; 97US-063542P.
XX 28-OCT-1997; 97US-063544P.
XX 28-OCT-1997; 97US-063549P.
XX 28-OCT-1997; 97US-063550P.
XX 28-OCT-1997; 97US-063564P.

```

PR 29-OCT-1997; 97US-063435P.
 PR 29-OCT-1997; 97US-063704P.
 PR 29-OCT-1997; 97US-063732P.
 PR 29-OCT-1997; 97US-063734P.
 PR 29-OCT-1997; 97US-063735P.
 PR 29-OCT-1997; 97US-063738P.
 PR 29-OCT-1997; 97US-064215P.
 PR 31-OCT-1997; 97US-063870P.
 PR 31-OCT-1997; 97US-064103P.
 PR 03-NOV-1997; 97US-064248P.
 PR 07-NOV-1997; 97US-064809P.
 PR 12-NOV-1997; 97US-065186P.
 PR 17-NOV-1997; 97US-065846P.
 PR 18-NOV-1997; 97US-065693P.
 PR 21-NOV-1997; 97US-066120P.
 PR 21-NOV-1997; 97US-066364P.
 PR 24-NOV-1997; 97US-066453P.
 PR 24-NOV-1997; 97US-066456P.
 PR 24-NOV-1997; 97US-066511P.
 PR 24-NOV-1997; 97US-066770P.
 PR 24-NOV-1997; 97US-066772P.
 PR 18-SEP-2000; 2000US-0665350.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi A, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
 PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
 PI Godowski PJ, Grimaldi JC, Gurney AL, Hillan KJ, Kljavin IJ;
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
 PI Williams PM, Wood WI;
 XX
 DR WPI; 2003-329602/31.
 DR N-PSDB; ACA60112.
 XX
 PT New transmembrane polypeptides and nucleic acids encoding the
 PT polypeptides, useful in gene therapy, in chromosome identification, as
 PT chromosome markers, in generating probes and in tissue typing
 XX
 PS Claim 12; Fig 50; 484pp; English.
 XX
 CC The invention relates to an isolated nucleic acid with at least 80%
 CC nucleic acid sequence identity to a nucleotide sequence encoding one of
 CC 61 secreted/transmembrane polypeptides, or PRO polypeptides or encoding a
 CC PRO protein extracellular domain. Also included are a vector comprising
 CC the PRO nucleic acid, a host cell comprising the vector, producing a PRO
 CC polypeptide (by culturing the host cell for the expression of the PRO
 CC polypeptide, and recovering the PRO polypeptide from the cell culture),
 CC an isolated PRO polypeptide (having at least 80% sequence identity
 CC to: (a) an amino acid sequence selected from the 61 PRO proteins;
 CC (b) an amino acid sequence encoded by a nucleic acid molecule deposited
 CC with an ATCC number (detailed in the specification); or (c) an
 CC extracellular domain of a PRO polypeptide or to a PRO polypeptide lacking
 CC its associated signal peptide), a chimeric molecule comprising a PRO
 CC polypeptide of fused to a heterologous amino acid sequence, an anti-PRO
 CC antibody, detecting a PRO245 or PRO1868 in a sample suspected of
 CC containing the polypeptide, linking a bioactive molecule to a cell
 CC expressing a PRO245 or PRO1868 and modulating at least one biological
 CC activity of a cell expressing a PRO245 or PRO1868. Nucleic acids which
 CC encode PRO can be used to generate either transgenic animals or knock-out
 CC animals which may be used in the development and screening of
 CC therapeutically useful reagents. The nucleic acids may also be used in
 CC gene therapy, in chromosome identification, as chromosome markers, or in
 CC generating probes. The PRO polypeptides are useful as molecular markers
 CC for protein electrophoresis, and the isolated nucleic acids may be used
 CC for recombinantly expressing those markers. The PRO polypeptides and
 CC nucleic acids may also be used in tissue typing. Anti-PRO antibodies
 CC are useful in diagnostic assays for PRO, and in affinity purification
 CC of PRO from recombinant cell culture or natural sources. The
 CC present sequence represents a PRO protein.

XX Sequence 382 AA;

Query Match 98.7%; Score 1101; DB 24; Length 382;

Best Local Similarity 96.3%; Pred. No. 9.6e-104;
 Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
 QY 1 ATGRLLS-----GQPVCRGTQPCYKVIYFHDTSRLNFEFEAKACRRDGGQLVSI 52
 Db |||||
 QY 22 ATGRLLSASDLRLRGQPVCRGTQPCYKVIYFHDTSRLNFEFEAKACRRDGGQLVSI 81
 Db |||||
 QY 53 ESEDEOKLIEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVD 112
 Db |||||
 QY 82 ESEDEOKLIEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVD 141
 Db |||||
 QY 113 EPCSGSEVCVMYHOPSPAGIGGPMFQWNNDRCKMKNFICKYSDKPAVPSRAEAGE 172
 Db |||||
 QY 142 EPCSGSEVCVMYHOPSPAGIGGPMFQWNNDRCKMKNFICKYSDKPAVPSRAEAGE 201
 Db |||||
 QY 173 ETELTPVLPPEETQBEDAKKTFKESREAAALNLAY 206
 Db |||||
 QY 202 ETELTPVLPPEETQBEDAKKTFKESREAAALNLAY 235
 Db |||||
 RESULT 14
 ABU65578
 ID ABU65578 standard; Protein; 382 AA.
 XX
 AC ABU65578;
 XX
 DT 19-MAY-2003 (first entry)
 XX
 DE Human secreted/transmembrane protein, SEQ ID 20.
 XX
 KW Human; PRO; secreted protein; transmembrane protein;
 KW cytostatic; antiarthritic; osteopathic; adrenal tumour; lung tumour;
 KW colon tumour; breast tumour; prostate tumour; rectal tumour;
 KW cervical tumour; liver tumour; TNF-alpha release; arthritis;
 KW tumour necrosis factor alpha; chondrocyte cell; bone disorder;
 KW cartilage disorder; sports injury.
 XX
 OS Homo sapiens.
 XX
 PN US2003036156-A1.
 PL 20-FEB-2003.
 XX
 PF 02-JUL-2002; 2002US-0188767.
 XX
 PR 16-SEP-1998; 98WO-US19330.
 PR 07-OCT-1998; 98WO-US21141.
 PR 01-DEC-1998; 98WO-US25108.
 PR 08-MAR-1999; 99WO-US05028.
 PR 14-MAY-1999; 99WO-US10733.
 PR 02-JUN-1999; 99WO-US12252.
 PR 01-SEP-1999; 99WO-US20111.
 PR 15-SEP-1999; 99WO-US21090.
 PR 01-DEC-1999; 99WO-US28301.
 PR 02-DEC-1999; 99WO-US28551.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05841.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.

PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 29-JUN-2001; 2001WO-US21066.
PR 09-JUL-2001; 2001WO-US21735.
PR 29-AUG-2001; 2001WO-US27099.
PR 18-SEP-1997; 97US-059263P.
PR 18-SEP-1997; 97US-059266P.
PR 17-OCT-1997; 97US-062250P.
PR 21-OCT-1997; 97US-063486P.
PR 24-OCT-1997; 97US-063120P.
PR 24-OCT-1997; 97US-063121P.
PR 28-OCT-1997; 97US-063540P.
PR 28-OCT-1997; 97US-063541P.
PR 28-OCT-1997; 97US-063544P.
PR 28-OCT-1997; 97US-063564P.
PR 29-OCT-1997; 97US-063734P.
PR 31-OCT-1997; 97US-063870P.
PR 31-OCT-1997; 97US-064103P.
PR 13-NOV-1997; 97US-065311P.
PR 21-NOV-1997; 97US-066120P.
PR 24-NOV-1997; 97US-066466P.
PR 24-NOV-1997; 97US-066772P.
PR 11-DEC-1997; 97US-069335P.
PR 12-DEC-1997; 97US-069435P.
PR 17-DEC-1997; 97US-069870P.
PR 18-DEC-1997; 97US-068017P.
PR 10-MAR-1998; 98US-077450P.
PR 11-MAR-1998; 98US-077632P.
PR 11-MAR-1998; 98US-077649P.
PR 20-MAR-1998; 98US-078886P.
PR 20-MAR-1998; 98US-078939P.
PR 27-MAR-1998; 98US-079664P.
PR 27-MAR-1998; 98US-079786P.
PR 31-MAR-1998; 98US-080107P.
PR 31-MAR-1998; 98US-080194P.
PR 01-APR-1998; 98US-080327P.
PR 01-APR-1998; 98US-080333P.
PR 08-APR-1998; 98US-081049P.
PR 08-APR-1998; 98US-081070P.
PR 09-APR-1998; 98US-081195P.
PR 15-APR-1998; 98US-081838P.
PR 21-APR-1998; 98US-082568P.
PR 21-APR-1998; 98US-082569P.
PR 22-APR-1998; 98US-082704P.
PR 22-APR-1998; 98US-082797P.
PR 28-APR-1998; 98US-083322P.
PR 29-APR-1998; 98US-083495P.
PR 29-APR-1998; 98US-083496P.
PR 29-APR-1998; 98US-083499P.
PR 29-APR-1998; 98US-083559P.
PR 05-MAY-1998; 98US-084366P.
PR 06-MAY-1998; 98US-084414P.
PR 07-MAY-1998; 98US-084639P.
PR 07-MAY-1998; 98US-084640P.
PR 07-MAY-1998; 98US-084643P.
PR 15-MAY-1998; 98US-085579P.
PR 15-MAY-1998; 98US-085580P.
PR 15-MAY-1998; 98US-085582P.
PR 18-MAY-1998; 98US-085700P.
PR 22-MAY-1998; 98US-086322P.
PR 22-MAY-1998; 98US-086486P.
PR 28-MAY-1998; 98US-087088P.
PR 28-MAY-1998; 98US-087098P.
PR 02-JUN-1998; 98US-087609P.
PR 03-JUN-1998; 98US-087759P.
PR 04-JUN-1998; 98US-087827P.
PR 04-JUN-1998; 98US-088025P.
PR 04-JUN-1998; 98US-088028P.
PR 04-JUN-1998; 98US-088029P.
PR 04-JUN-1998; 98US-088033P.
PR 04-JUN-1998; 98US-088326P.
PR 05-JUN-1998; 98US-088167P.
PR 05-JUN-1998; 98US-088202P.
PR 05-JUN-1998; 98US-088212P.
PR 05-JUN-1998; 98US-088217P.
PR 05-JUN-1998; 98US-088655P.
PR 10-JUN-1998; 98US-088722P.
PR 10-JUN-1998; 98US-088738P.
PR 10-JUN-1998; 98US-088740P.
PR 10-JUN-1998; 98US-088811P.
PR 10-JUN-1998; 98US-088824P.
PR 10-JUN-1998; 98US-088825P.
PR 10-JUN-1998; 98US-088826P.
PR 11-JUN-1998; 98US-088861P.
PR 11-JUN-1998; 98US-088863P.
PR 11-JUN-1998; 98US-088876P.
PR 12-JUN-1998; 98US-089090P.
PR 12-JUN-1998; 98US-089105P.
PR 16-JUN-1998; 98US-089512P.
PR 16-JUN-1998; 98US-089514P.
PR 17-JUN-1998; 98US-089538P.
PR 17-JUN-1998; 98US-089598P.
PR 17-JUN-1998; 98US-089653P.
PR 18-JUN-1998; 98US-089908P.
PR 19-JUN-1998; 98US-089952P.
PR 22-JUN-1998; 98US-090246P.
PR 22-JUN-1998; 98US-090252P.
PR 24-JUN-1998; 98US-090254P.
PR 24-JUN-1998; 98US-090429P.
PR 24-JUN-1998; 98US-090435P.
PR 24-JUN-1998; 98US-090444P.
PR 24-JUN-1998; 98US-090461P.
PR 24-JUN-1998; 98US-090535P.
PR 24-JUN-1998; 98US-090540P.
PR 25-JUN-1998; 98US-090676P.
PR 25-JUN-1998; 98US-090678P.
PR 25-JUN-1998; 98US-090688P.
PR 25-JUN-1998; 98US-090690P.
PR 25-JUN-1998; 98US-090694P.
PR 25-JUN-1998; 98US-090695P.
PR 25-JUN-1998; 98US-090896P.
PR 26-JUN-1998; 98US-090862P.
PR 26-JUN-1998; 98US-090863P.
PR 26-JUN-1998; 98US-091010P.
PR 01-JUL-1998; 98US-091359P.
PR 01-JUL-1998; 98US-091544P.
PR 02-JUL-1998; 98US-091478P.
PR 02-JUL-1998; 98US-091486P.
PR 02-JUL-1998; 98US-091626P.
PR 02-JUL-1998; 98US-091628P.
PR 02-JUL-1998; 98US-091632P.
PR 24-JUL-1998; 98US-094006P.
PR 04-AUG-1998; 98US-095282P.
PR 10-AUG-1998; 98US-095998P.
PR 10-AUG-1998; 98US-096012P.
PR 17-AUG-1998; 98US-096757P.
PR 17-AUG-1998; 98US-096766P.
PR 17-AUG-1998; 98US-096867P.
PR 17-AUG-1998; 98US-096891P.
PR 17-AUG-1998; 98US-096897P.
PR 18-AUG-1998; 98US-096949P.
PR 18-AUG-1998; 98US-096959P.
PR 26-AUG-1998; 98US-097022P.
PR 26-AUG-1998; 98US-097522P.
PR 26-AUG-1998; 98US-097954P.
PR 26-AUG-1998; 98US-097955P.
PR 26-AUG-1998; 98US-097971P.
PR 26-AUG-1998; 98US-097974P.
PR 01-SEP-1998; 98US-098014P.
PR 01-SEP-1998; 98US-098716P.
PR 02-SEP-1998; 98US-098723P.
PR 02-SEP-1998; 98US-098803P.
PR 02-SEP-1998; 98US-098821P.

```
PR 02-SEP-1998; 98US-098843P.
PR 09-SEP-1998; 98US-099602P.
PR 10-SEP-1998; 98US-099741P.

Query Match      98.7%; Score 1101; DB 24; Length 382;
Best Local Similarity 96.3%; Pred. No. 9.6e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSI 52
Db 22 ATGRLLSASDLDRGGQVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSI 81

Qy 53 ESEDEQKLIKFIEIENLPSDGFWMIGLRRRBEKQSNSTACQDLVATWDGSIQFRNMYVD 112
Db 82 ESEDEQKLIKFIEIENLPSDGFWMIGLRRRBEKQSNSTACQDLVATWDGSIQFRNMYVD 141

Qy 113 EPSCGSEVCVVMYHQPSPAGIGGYPMEQWDDRCNMKNFICKYSDEKPAVPSREAGE 172
Db 142 EPSCGSEVCVVMYHQPSPAGIGGYPMEQWDDRCNMKNFICKYSDEKPAVPSREAGE 201

Qy 173 ETELTTPVLPETQEDAKTKFKESREAAALNLAY 206
Db 202 ETELTTPVLPETQEDAKTKFKESREAAALNLAY 235

RESULT 15
ABU65911
ID ABU65911 standard; Protein; 382 AA.
AC ABU65911;
XX
XX
XX 20-MAY-2003 (first entry)
XX
XX Novel human secreted and transmembrane protein PRO234.
XX
XX Human; secreted protein; transmembrane protein; cytostatic;
KW gene therapy; TNF-Agonist-Alpha; chondrocyte stimulator; tumour;
KW adrenal tumour; lung tumour; colon tumour; breast tumour;
KW prostate tumour; rectal tumour; cervical tumour; liver tumour.
XX
XX Homo sapiens.
XX
XX US2003036157-A1.
XX
XX 20-FEB-2003.
XX
XX 02-JUL-2002; 2002US-0188769.
XX
XX 16-SEP-1998; 98WO-US19330.
XX 07-OCT-1998; 98WO-US21141.
XX 01-DEC-1998; 98WO-US25108.
XX 08-MAR-1999; 99WO-US05028.
XX 14-MAY-1999; 99WO-US10733.
XX 02-JUN-1999; 99WO-US12252.
XX 01-SEP-1999; 99WO-US20111.
XX 15-SEP-1999; 99WO-US21090.
XX 01-DEC-1999; 99WO-US28301.
XX 02-DEC-1999; 99WO-US28551.
XX 30-DEC-1999; 99WO-US31274.
XX 05-JAN-2000; 2000WO-US00219.
XX 18-FEB-2000; 2000WO-US04341.
XX 18-FEB-2000; 2000WO-US04342.
XX 22-FEB-2000; 2000WO-US04414.
XX 24-FEB-2000; 2000WO-US05004.
XX 01-MAR-2000; 2000WO-US05601.
XX 02-MAR-2000; 2000WO-US05841.
XX 15-MAR-2000; 2000WO-US06884.
XX 30-MAR-2000; 2000WO-US08439.
XX 17-MAY-2000; 2000WO-US13705.
XX 22-MAY-2000; 2000WO-US14042.
XX 30-MAY-2000; 2000WO-US14941.
XX 02-JUN-2000; 2000WO-US15264.
XX 28-JUL-2000; 2000WO-US20710.

24-AUG-2000; 2000WO-US23328.
08-NOV-2000; 2000WO-US30952.
01-DEC-2000; 2000WO-US32678.
20-DEC-2000; 2000WO-US34956.
28-FEB-2001; 2001WO-US06520.
01-JUN-2001; 2001WO-US17800.
20-JUN-2001; 2001WO-US19692.
29-JUN-2001; 2001WO-US21066.
09-JUL-2001; 2001WO-US21735.
29-AUG-2001; 2001WO-US27099.
18-SEP-1997; 97US-059263P.
18-SEP-1997; 97US-059266P.
17-OCT-1997; 97US-062250P.
21-OCT-1997; 97US-063486P.
24-OCT-1997; 97US-063120P.
24-OCT-1997; 97US-063121P.
28-OCT-1997; 97US-063540P.
28-OCT-1997; 97US-063541P.
28-OCT-1997; 97US-063544P.
28-OCT-1997; 97US-063564P.
29-OCT-1997; 97US-063734P.
31-OCT-1997; 97US-063870P.
31-OCT-1997; 97US-064103P.
13-NOV-1997; 97US-065311P.
21-NOV-1997; 97US-066120P.
24-NOV-1997; 97US-066466P.
11-DEC-1997; 97US-066772P.
29-OCT-1997; 97US-069335P.
12-DEC-1997; 97US-069425P.
17-DEC-1997; 97US-069870P.
18-DEC-1997; 97US-068017P.
11-MAR-1998; 98US-077450P.
11-MAR-1998; 98US-077632P.
11-MAR-1998; 98US-077649P.
20-MAR-1998; 98US-078886P.
20-MAR-1998; 98US-078939P.
27-MAR-1998; 98US-079664P.
27-MAR-1998; 98US-079786P.
31-MAR-1998; 98US-080107P.
31-MAR-1998; 98US-080194P.
01-APR-1998; 98US-080327P.
01-APR-1998; 98US-080333P.
08-APR-1998; 98US-081049P.
08-APR-1998; 98US-081070P.
09-APR-1998; 98US-081195P.
15-APR-1998; 98US-081838P.
21-APR-1998; 98US-082568P.
21-APR-1998; 98US-082569P.
22-APR-1998; 98US-082704P.
22-APR-1998; 98US-082797P.
28-APR-1998; 98US-083322P.
29-APR-1998; 98US-083495P.
29-APR-1998; 98US-083496P.
29-APR-1998; 98US-083499P.
29-APR-1998; 98US-083559P.
05-MAY-1998; 98US-084366P.
06-MAY-1998; 98US-084414P.
07-MAY-1998; 98US-084639P.
07-MAY-1998; 98US-084640P.
07-MAY-1998; 98US-084643P.
15-MAY-1998; 98US-085579P.
15-MAY-1998; 98US-085580P.
15-MAY-1998; 98US-085582P.
15-MAY-1998; 98US-085700P.
18-MAY-1998; 98US-086023P.
22-MAY-1998; 98US-086392P.
22-MAY-1998; 98US-086486P.
28-MAY-1998; 98US-087098P.
28-MAY-1998; 98US-087208P.
02-JUN-1998; 98US-087609P.
02-JUN-1998; 98US-087759P.
03-JUN-1998; 98US-087827P.
04-JUN-1998; 98US-088025P.
```

| | | |
|----|--------------|----------------|
| PR | 04-JUN-1998; | 98US-08802821P |
| PR | 04-JUN-1998; | 98US-0880299P |
| PR | 04-JUN-1998; | 98US-0880333P |
| PR | 04-JUN-1998; | 98US-0880359P |
| PR | 05-JUN-1998; | 98US-0881266P |
| PR | 05-JUN-1998; | 98US-0881676P |
| PR | 05-JUN-1998; | 98US-088202P |
| PR | 05-JUN-1998; | 98US-088202P |
| PR | 05-JUN-1998; | 98US-088212P |
| PR | 05-JUN-1998; | 98US-088217P |
| PR | 05-JUN-1998; | 98US-088555P |
| PR | 10-JUN-1998; | 98US-088722P |
| PR | 10-JUN-1998; | 98US-088738P |
| PR | 10-JUN-1998; | 98US-088740P |
| PR | 10-JUN-1998; | 98US-088811P |
| PR | 10-JUN-1998; | 98US-088824P |
| PR | 10-JUN-1998; | 98US-088825P |
| PR | 10-JUN-1998; | 98US-088826P |
| PR | 11-JUN-1998; | 98US-088861P |
| PR | 11-JUN-1998; | 98US-088863P |
| PR | 11-JUN-1998; | 98US-088867P |
| PR | 12-JUN-1998; | 98US-088909P |
| PR | 12-JUN-1998; | 98US-089105P |
| PR | 16-JUN-1998; | 98US-089312P |
| PR | 16-JUN-1998; | 98US-089314P |
| PR | 17-JUN-1998; | 98US-089538P |
| PR | 17-JUN-1998; | 98US-089598P |
| PR | 17-JUN-1998; | 98US-089653P |
| PR | 18-JUN-1998; | 98US-089908P |
| PR | 19-JUN-1998; | 98US-089952P |
| PR | 22-JUN-1998; | 98US-090246P |
| PR | 22-JUN-1998; | 98US-090252P |
| PR | 22-JUN-1998; | 98US-090254P |
| PR | 24-JUN-1998; | 98US-090439P |
| PR | 24-JUN-1998; | 98US-090443P |
| PR | 24-JUN-1998; | 98US-090446P |
| PR | 24-JUN-1998; | 98US-090461P |
| PR | 24-JUN-1998; | 98US-090535P |
| PR | 24-JUN-1998; | 98US-090540P |
| PR | 25-JUN-1998; | 98US-090676P |
| PR | 25-JUN-1998; | 98US-090678P |
| PR | 25-JUN-1998; | 98US-090688P |
| PR | 25-JUN-1998; | 98US-090690P |
| PR | 25-JUN-1998; | 98US-090694P |
| PR | 25-JUN-1998; | 98US-090695P |
| PR | 25-JUN-1998; | 98US-090696P |
| PR | 26-JUN-1998; | 98US-090862P |
| PR | 26-JUN-1998; | 98US-090863P |
| PR | 26-JUN-1998; | 98US-091010P |
| PR | 01-JUL-1998; | 98US-091135P |
| PR | 02-JUL-1998; | 98US-091544P |
| PR | 02-JUL-1998; | 98US-091478P |
| PR | 02-JUL-1998; | 98US-091486P |
| PR | 02-JUL-1998; | 98US-091626P |
| PR | 02-JUL-1998; | 98US-091628P |
| PR | 02-JUL-1998; | 98US-091632P |
| PR | 24-JUL-1998; | 98US-094006P |
| PR | 04-AUG-1998; | 98US-095282P |
| PR | 10-AUG-1998; | 98US-095998P |
| PR | 10-AUG-1998; | 98US-096012P |
| PR | 17-AUG-1998; | 98US-096757P |
| PR | 17-AUG-1998; | 98US-096766P |
| PR | 17-AUG-1998; | 98US-096867P |
| PR | 17-AUG-1998; | 98US-096891P |
| PR | 18-AUG-1998; | 98US-096897P |
| PR | 18-AUG-1998; | 98US-096949P |
| PR | 18-AUG-1998; | 98US-096959P |
| PR | 25-AUG-1998; | 98US-097022P |
| PR | 26-AUG-1998; | 98US-097952P |
| PR | 26-AUG-1998; | 98US-097954P |
| PR | 26-AUG-1998; | 98US-097956P |
| PR | 26-AUG-1998; | 98US-097971P |
| PR | 26-AUG-1998; | 98US-097974P |
| PR | 26-AUG-1998; | 98US-098014P |
| PR | 01-SEP-1998; | 98US-098716P |

| | | |
|----|--------------|---------------|
| PR | 01-SEP-1998; | 98US-098723P; |
| PR | 02-SEP-1998; | 98US-098803P; |
| PR | 02-SEP-1998; | 98US-098821P; |
| PR | 02-SEP-1998; | 98US-098843P; |
| PR | 09-SEP-1998; | 98US-098602P; |
| PR | 10-SEP-1998; | 98US-098741P; |
| PR | 10-SEP-1998; | 98US-098755P; |
| PR | 10-SEP-1998; | 98US-098763P; |

Query Match 98.7%; Score 1101; DB 24; Length 382;
Best Local Similarity 96.3%; Pred. No. 9.6e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 8

| | | | |
|----|-----|---|-----|
| QY | 1 | ATGRLLS-----GQPCVCGGTQRCYKVIYFHDTSRRLNFEEAKEACRRGGQLVSI | 52 |
| Db | 22 | ATGRLLSASDLDRGGQPCVCGGTQRCYKVIYFHDTSRRLNFEEAKEACRRGGQLVSI | 81 |
| QY | 53 | ESEDEQKLEIKFIENLLPSDGDWFIGLRRREEKQSNSTACQDLYAWTDGSI | 112 |
| Db | 82 | ESEDEQKLEIKFIENLLPSDGDWFIGLRRREEKQSNSTACQDLYAWTDGSI | 141 |
| QY | 113 | EPSCGSEVCVMYHQPSAPAGIGGYPYFQWDDRCNMKNNFICKYDCKPAVPSREAAGE | 172 |
| Db | 142 | EPSCGSEVCVMYHQPSAPAGIGGYPYFQWDDRCNMKNNFICKYDCKPAVPSREAAGE | 201 |
| QY | 173 | ETELTTPVLPEETOEDAKKTFKESREAAALNAY | 206 |
| Db | 202 | ETELTTPVLPEETOEDAKKTFKESREAAALNAY | 235 |

Search completed: December 22, 2003, 16:10:09
Job time : 29.1819 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:05:39 ; Search time 24.0017 Seconds
(without alignments)
2214.797 Million cell updates/sec

Title: US-09-887-855-2_COPY_22_227

Perfect score: 1115
Sequence: 1 ATGRLSGQPCVCRGGTQRPC.....EEDAKTFKESREAAALNLAY 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mmc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_undefined.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|---------------------|
| 1 | 1115 | 100.0 | 374 | Q8TAY8 | Q8TAY8 homo sapien |
| 2 | 1115 | 100.0 | 374 | Q96NF3 | Q96NF3 homo sapien |
| 3 | 1111 | 99.6 | 374 | Q96NC5 | Q96NC5 homo sapien |
| 4 | 945.5 | 84.8 | 374 | Q92209 | Q92209 cricetus |
| 5 | 803.5 | 72.1 | 211 | Q8C351 | Q8C351 mus musculus |
| 6 | 580 | 52.0 | 246 | Q8BVI7 | Q8BVI7 mus musculus |
| 7 | 580 | 52.0 | 292 | Q8BVU2 | Q8BVU2 mus musculus |
| 8 | 185 | 16.6 | 1290 | Q9W6E1 | Q9W6E1 gallus gall |
| 9 | 178.5 | 16.0 | 1456 | Q61830 | Q61830 mus musculus |
| 10 | 177.5 | 15.9 | 1348 | Q25199 | Q25199 hydra atten |
| 11 | 176.5 | 15.8 | 1479 | Q9Y5P9 | Q9Y5P9 homo sapien |
| 12 | 176.5 | 15.8 | 1479 | Q9UBG0 | Q9UBG0 homo sapien |
| 13 | 170.5 | 15.3 | 1479 | Q64449 | Q64449 mus musculus |
| 14 | 169 | 15.2 | 217 | Q8C4F8 | Q8C4F8 mus musculus |
| 15 | 165.5 | 14.8 | 742 | Q8K4Q8 | Q8K4Q8 mus musculus |
| 16 | 165.5 | 14.8 | 742 | Q8C979 | Q8C979 mus musculus |

| | | | | | | |
|----|-------|------|------|----|--------|---------------------|
| 17 | 161.5 | 14.5 | 134 | 5 | Q9XXY3 | Q9XXY3 hydra magni |
| 18 | 159.5 | 14.3 | 742 | 11 | Q8VIF6 | Q8VIF6 mus musculus |
| 19 | 158.5 | 14.2 | 719 | 6 | O62623 | O62623 bos taurus |
| 20 | 158 | 14.2 | 142 | 11 | Q8CJ86 | Q8CJ86 mus musculus |
| 21 | 158 | 14.2 | 142 | 11 | Q8BHK7 | Q8BHK7 mus musculus |
| 22 | 158 | 14.2 | 295 | 11 | Q912W4 | Q912W4 mus musculus |
| 23 | 158 | 14.2 | 311 | 11 | Q9D8V4 | Q9D8V4 mus musculus |
| 24 | 158 | 14.2 | 325 | 11 | Q912X0 | Q912X0 mus musculus |
| 25 | 157 | 14.1 | 158 | 13 | Q90W17 | Q90W17 bungarus fa |
| 26 | 156 | 14.0 | 323 | 11 | Q8CJ91 | Q8CJ91 mus musculus |
| 27 | 156 | 14.0 | 339 | 6 | Q95244 | Q95244 sus scrofa |
| 28 | 155 | 13.9 | 293 | 11 | Q8BG20 | Q8BG20 mus musculus |
| 29 | 155 | 13.9 | 323 | 11 | Q8CJ94 | Q8CJ94 mus musculus |
| 30 | 155 | 13.9 | 323 | 11 | Q8CJ93 | Q8CJ93 mus musculus |
| 31 | 155 | 13.9 | 323 | 11 | Q8CJ88 | Q8CJ88 mus musculus |
| 32 | 155 | 13.9 | 1152 | 13 | Q90WM2 | Q90WM2 xenopus lae |
| 33 | 154 | 13.8 | 322 | 11 | Q8CJ89 | Q8CJ89 mus musculus |
| 34 | 154 | 13.8 | 323 | 11 | Q8CJ92 | Q8CJ92 mus musculus |
| 35 | 152.5 | 13.7 | 485 | 6 | Q95LG3 | Q95LG3 odocoileus |
| 36 | 152 | 13.6 | 158 | 13 | Q90W16 | Q90W16 bungarus mu |
| 37 | 151.5 | 13.6 | 652 | 4 | Q8IXK1 | Q8IXK1 homo sapien |
| 38 | 150.5 | 13.5 | 399 | 6 | Q8HY12 | Q8HY12 hylobates l |
| 39 | 150 | 13.5 | 197 | 6 | Q28008 | Q28008 bos taurus |
| 40 | 149.5 | 13.4 | 158 | 11 | Q8JZX6 | Q8JZX6 mus musculus |
| 41 | 149.5 | 13.4 | 459 | 5 | Q22136 | Q22136 caenorhabdi |
| 42 | 149 | 13.4 | 158 | 13 | Q90W18 | Q90W18 bungarus fa |
| 43 | 148.5 | 13.3 | 163 | 13 | Q8AXR8 | Q8AXR8 anguilla ja |
| 44 | 148.5 | 13.3 | 195 | 5 | Q27340 | Q27340 megabalanus |
| 45 | 148.5 | 13.3 | 330 | 4 | Q9NT67 | Q9NT67 homo sapien |

ALIGNMENTS

RESULT 1

Q8TAY8
ID Q8TAY8 PRELIMINARY; PRT; 374 AA.
AC Q8TAY8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
D 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Similar to unnamed protein product.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025407; AAH25407.1; -
DR InterPro; IPR001304; Lectin_C;
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00041; C TYPE LECTIN 2; 1.
SQ SEQUENCE 374 AA; -42312 MW; FC214E6BC9E578D9 CRC64;

Query Match 100.0%; Score 1115; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. NO. 4.5e-99;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 1 | ATGRLLSGQPCVCRGGTQRPCVYIYFHDTSRRLNFEAKACRRDGGQLVSI | 60 |
| Db | 22 | ATGRLLSGQPCVCRGGTQRPCVYIYFHDTSRRLNFEAKACRRDGGQLVSI | 81 |
| QY | 61 | IEKFIENLPSDGFWIGLRRREEKQSNSTACQDLYAWTDGSI | 120 |
| Db | 82 | IEKFIENLPSDGFWIGLRRREEKQSNSTACQDLYAWTDGSI | 141 |
| QY | 121 | CVVMYHQPSAPAGIGGPFQWNNDRCNMKNFICKYSDKPAVPSREAGE | 180 |
| Db | 142 | CVVMYHQPSAPAGIGGPFQWNNDRCNMKNFICKYSDKPAVPSREAGE | 201 |


```

QY 181 LPBETQEDAKTKFKESREALNLAY 206
DB 202 LPBETQEDAKTKFKESREALNLAY 227

RESULT 2
Q96NF3 ID Q96NF3 PRELIMINARY; PRT; 374 AA.
AC Q96NF3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ30977.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotuta T., Hirooka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsu K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RE EMBL; AK055339; BAB70946.1; -.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin.c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 374 AA; 42280 MW; 8AE64E6BC9E56DCD CRC64;

Query Match 100.0%; Score 1115; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 4.5e-99;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSIIESEDEQKL 60
DB 22 ATGRLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSIIESEDEQKL 81
QY 61 IEKFIENLLPSDGDGFWIGLRREEKQSNSTACQDLYAWTDGSIQSRNMYVDPSGCGSEV 120
DB 82 IEKFIENLLPSDGDGFWIGLRREEKQSNSTACQDLYAWTDGSIQSRNMYVDPSGCGSEV 141
QY 121 CVVMYHQPAPAGIGGYPYFQWDDRCNKNFNICKYSDKPAVPSREAGEETELTPV 180
DB 142 CVVMYHQPAPAGIGGYPYFQWDDRCNKNFNICKYSDKPAVPSREAGEETELTPV 201
QY 181 LPBETQEDAKTKFKESREALNLAY 206
DB 202 LPBETQEDAKTKFKESREALNLAY 227

RESULT 3
Q96NC5 ID Q96NC5 PRELIMINARY; PRT; 374 AA.
AC Q96NC5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ31092.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Taishiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsu K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RE EMBL; AK055654; BAB70978.1; -.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin.c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 374 AA; 42310 MW; CBF74E676E23BA16 CRC64;

Query Match 99.6%; Score 1111; DB 4; Length 374;
Best Local Similarity 99.5%; Pred. No. 1.1e-98;
Matches 205; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGRLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSIIESEDEQKL 60
DB 22 ATGRLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSIIESEDEQKL 81
QY 61 IEKFIENLLPSDGDGFWIGLRREEKQSNSTACQDLYAWTDGSIQSRNMYVDPSGCGSEV 120
DB 82 IEKFIENLLPSDGDGFWIGLRREEKQSNSTACQDLYAWTDGSIQSRNMYVDPSGCGSEV 141
QY 121 CVVMYHQPAPAGIGGYPYFQWDDRCNKNFNICKYSDKPAVPSREAGEETELTPV 180
DB 142 CVVMYHQPAPAGIGGYPYFQWDDRCNKNFNICKYSDKPAVPSREAGEETELTPV 201
QY 181 LPBETQEDAKTKFKESREALNLAY 206
DB 202 LPBETQEDAKTKFKESREALNLAY 227

RESULT 4
Q92209 ID Q92209 PRELIMINARY; PRT; 374 AA.
AC Q92209;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Layilin.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Borowsky M.L., Hynes R.O.;
RT "Layilin, a novel talin-binding transmembrane protein homologous with
RT C-type lectins, is localized in membrane ruffles.";
RL J. Cell Biol. 143:0-0(1998).
DR EMBL; AF093673; AAC68695.1; -.
DR HSSP; P22897; 1EGG.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin.c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
SQ SEQUENCE 374 AA; 42435 MW; 298A8BA24F04E1C CRC64;

Query Match 84.8%; Score 945.5; DB 11; Length 374;
Best Local Similarity 85.9%; Pred. No. 9.4e-83;
Matches 176; Conservative 9; Mismatches 19; Indels 1; Gaps 1;

QY 3 GRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSIIESEDEQKL 62
DB 24 GRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSIIESEDEQKL 83

```


Db 197 NQ--PEETHE 204
|||||

RESULT 8

Q9W6E1 ID Q9W6E1 PRELIMINARY; PRT; 1290 AA.
AC Q9W6E1; DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Neurocan core protein.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20309833; PubMed=10851024;
RA Li H., Leung T.C., Hoffman S., Baleamo J., Lillian J.;
RT "Coordinate Regulation of Cadherin and Integrin Function by the
RT Chondroitin Sulfate Proteoglycan Neurocan."
RL J. Cell Biol. 149:1275-1288 (2000).
DR EMBL; AF116856; AAD24546.2; -.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR00152; Asx hydroxyl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; IG; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR00010; EGFELOOD.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS50835; IG LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW EGF-like domain.
SQ SEQUENCE 1290 AA; 139877 MW; 182BD86D0E40BE78 CRC64;

Query Match 16.6%; Score 185; DB 13; Length 1290;
Best Local Similarity 34.7%; Pred. No. 7e-09;
Matches 50; Conservative 18; Mismatches 42; Indels 34; Gaps 8;
Qy 17 QRCVKVLYFHDTSRLNFEAKEACRDGGQGVLSIESDEQKLEKFIENLLPSDGF 76
Db 1064 QGHCVY--YF---SRRSWEAEDCRRAAGLTSIHSEHGFINSP-----GHENTW 1112
Qy 77 IGLRRREKQSNSTACQDLYAWTSGISQFRNYYVDEPS---CGSEVCVMY-HQPSAPA 132
Db 1113 IGLNDRIVEQD-----FQNTDTGLQYENWRENQDPNFAGDCDVLVSHE----- 1159

Qy 133 GIGGPFYMFQWDDRCNMKNPFICK 156
Db 1160 -IG-----KWNVDVPCYNLPYICK 1177
|||||

RESULT 9

Q61830 ID Q61830 PRELIMINARY; PRT; 1456 AA.
AC Q61830; DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Macrophage mannose receptor precursor.
GN MRC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6;
RX MEDLINE=93043353; PubMed=1421407;
RA Harris N., Rits M., Chang G., Ezekowitz R.B.;
RT "Characterization of the murine macrophage mannose receptor."
RL Blood 80:2363-2373 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6;
RA Super M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z11974; CAA78028.1; -.
DR HSSP; P22897; 1EGG.
DR MGD; MGI:971142; Mrc1.
DR InterPro; IPR002353; AntifreezeII.
DR InterPro; IPR000562; FN Type_II.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin_c; 8.
DR Pfam; PF00652; Ricin_B_lectin; 2.
DR PRINTS; PR00356; ANTIFREEZEII.
DR PRINTS; PR00013; FNTYPEII.
DR ProDom; PD000995; FN Type_II; 1.
DR SMART; SM00034; CLECT; 8.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 6.
DR PROSITE; PS50041; C-TYPE LECTIN_2; 8.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS50231; RICIN_B_LECTIN; 1.
KW Receptor; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 1456 MACROPHAGE MANNOSE RECEPTOR.
SQ SEQUENCE 1456 AA; 165065 MW; 4EBD3F1B8619A594 CRC64;

Query Match 16.0%; Score 178.5; DB 11; Length 1456;
Best Local Similarity 25.9%; Pred. No. 3.4e-08;
Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;
Qy 21 YKVIYFHDTSRLNFEAKEACRDGGQGVLSIESDEQKLEKFIENLLPSDGF 80
Db 807 YKDYQYFSEKKEKTDNARRCKNFGDLATIKSEKCKFLWKYI-NKNGQSPYFIGML 865
Qy 81 RREKQSNSTACQDLYAWTSGISQFRNYYVDEPSGSGS--EVCVMYHQPSAPAGIGPY 138
Db 866 ISMDKK-----FIWMDGSKYDFVAWATGEFNFANDDENCVTWY-----TNSGF---- 908
Qy 139 MFQWDDRCNMKNPFICK---YSDEKPAVPSREAGEETELTTPVLPEETQE----- 187
Db 909 ---WINDGCVPNFIQRIHNSINATAMP-----TTPTTGGCKEGHLYKNK 954
Qy 188 -----EDAKTTFKESREAAALNL 204
: ||:::|

```

Db      955 CFKIFGFANEKKSNQDARQACKGL 979

RESULT 10
Q25199 ID Q25199 PRELIMINARY; PRT; 1348 AA.
AC Q25199;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Tyrosine kinase receptor.
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6087;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Irvine;
RA MEDLINE=20209407; PubMed=10744720;
RX Reidling J.C., Miller M.A., Steele R.E.;
RT "Sweet Tooth, a Novel Receptor Protein-Tyrosine Kinase with C-type
RT Lectin-like Extracellular Domains.";
RL J. Biol. Chem. 275:10323-10330(2000).
DR EMBL; L22612; AAA29218.2; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR003990; Pncreatase_ac.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR01504; PNCREATITSAP.
DR ProDom; PD000001; Prot_kinase; 2.
DR SMART; SM00034; CLECT; 4.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 2.
DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1348 AA; 156916 MW; 77D2122093227FEF CRC64;

Query Match 15.9%; Score 177.5; DB 5; Length 1348;
Best Local Similarity 28.0%; Pred. No. 3.9e-08;
Matches 47; Conservative 32; Mismatches 46; Indels 43; Gaps 8;

QY 6 LSGQVPC--RGGTQRPC-----YKVIYFHDTSRLNPFEEAKEACRRDGGOLVSISEDE 57
Db 412 LSHRFICKVRATNEYCAEGWTSYRIYCFIYSIEFDWFKSFSCQNGNLLSIENQEE 471
QY 58 QKLEKTIENLLPSDGD-FWIGLRR-----REKQSNSTACQDLYAWTDGSIQFRNRY 110
Db 472 ----NRFIENDLIKNDKYIWLGNKINWYLNKKNR-----FEWSDNTYTFQFNWI 518
QY 111 VDEP--SCGSEVVCVMYHQPSAPAGIGGPMFQWDDRCNMKNFKICK 156
Db 519 TNQPDNNNGIESCEVMNIN-----GMSDKCKVLNGFICK 553

RESULT 11
Q9Y5P9 ID Q9Y5P9 PRELIMINARY; PRT; 1479 AA.
AC Q9Y5P9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Endocytic receptor Endo180.
GN ENDO180.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20148849; PubMed=10683150;
RA Sheikh H., Yarwood H., Ashworth A., Isacke C.;
RT "Endo180, an endocytic recycling glycoprotein related to the
RT cells and macrophages and functions as a lectin receptor.";
RL J. Cell Sci. 113:1021-1032(2000).
DR EMBL; AF134838; AAD30280.1; -.
DR HSSP; P02751; 2FN2.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000566; Lipocln_cytFABP.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin_c; 8.
DR PRINTS; PR00013; FNTYPEII.
DR ProDom; PD000995; FN_Type_II; 1.
DR SMART; SM00034; CLECT; 8.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 3.
DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 8.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS00213; LIPOCALIN; 1.
DR PROSITE; PS02031; RICIN_B_LLECTIN; 1.
KW Receptor.
SQ SEQUENCE 1479 AA; 166669 MW; 9F4BAF355F036FCE CRC64;

Query Match 15.8%; Score 176.5; DB 4; Length 1479;
Best Local Similarity 30.4%; Pred. No. 5.4e-08;
Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;

QY 7 SGQPVCRGGTQPCVKVIYFHDTSRLNPFEEAKEACRRDGGOLVSISEDEKLEKIE 66
Db 385 SWQPF-----QGHCTRL-----QAEKRSWQESKACLRGGDLVSIHMAELEFTTK 434
QY 67 NLPSPDGFWMIGLRRREEKQSNSTACQDLYAWTDGSIQFRNRYVDPS---CGSEVCW 123
Db 435 QEVE--ELMIGL-----NDLKLQWPFWSGSLVSVFTHHPFENNFRDLSLDCVT 483
QY 124 MYHPSAPAGIGGPMFQWDDRCNMKNFKICKYSDKPAVPSREAG 171
Db 484 IW-----GPEG-----RWNDSPCNQSLPSICKKAGQLSQGAABEDHG 520

RESULT 12
Q9UBG0 ID Q9UBG0 PRELIMINARY; PRT; 1479 AA.
AC Q9UBG0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Urokinase receptor-associated protein UPARAP.
GN KIAA0709.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can

```

| | | | |
|------------|--|--|--|
| RT | code for large proteins in vitro.; | | |
| RL | DNA Reb. 5:169-176(1998). | | |
| DR | EMBL; AF107232; AAF14192.1; - | | |
| DR | EMBL; AB014609; BAA31684.1; - | | |
| DR | HSSP; P02751; 2FN2. | | |
| DR | InterPro; IPR000562; FN Type II. | | |
| DR | InterPro; IPR001304; Lectin_C. | | |
| DR | InterPro; IPR000566; Lipoclin_CytfABP. | | |
| DR | InterPro; IPR000772; Ricin_B_lectin. | | |
| DR | Pfam; PF00040; fn2; 1. | | |
| DR | Pfam; PF00059; lectin.c; 8. | | |
| DR | PRINTS; PR00013; FNTYPEII. | | |
| DR | ProDom; PD000995; FN Type_II; 1. | | |
| DR | SMART; SM00034; CLECT; 8. | | |
| DR | SMART; SM00059; FN2; 1. | | |
| DR | SMART; SM00458; RICIN; 1. | | |
| DR | PROSITE; PS00615; C_TYPE_LECTIN_1; 3. | | |
| DR | PROSITE; PS00041; C_TYPE_LECTIN_2; 8. | | |
| DR | PROSITE; PS00023; FIBRONECTIN_2; 1. | | |
| DR | PROSITE; PS00213; LIPOCALIN; 1. | | |
| DR | PROSITE; PS02031; RICIN_B_LECTIN; 1. | | |
| KW | Kinase; Receptor. | | |
| SQ | SEQUENCE 1479 AA; 166654 MW; C7583EA78E2792D1 CRC64; | | |
| | Query Match | 15.8%; Score 176.5; DB 4; Length 1479; | |
| | Best Local Similarity | 30.4%; Pred. No. 5.4e-08; | |
| | Matches | 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7 | |
| Qy | 7 | SGQPVRCGGTGPCKVYIFHDTSRRLNFEAKEACRRDGGQLVSISEDEQKLIKPIE 66 | |
| Db | 385 | SWQPF-----QGHCVRL-----QAEKRSQWSEKSKACLGGGDLVSIHGMASLEFITTQIK 434 | |
| Qy | 67 | NLLPSDGFDTGLRRREKQNSSTACQDLVAWTGSIISQFRNWTYDEPS---CGSEVCVV 123 | |
| Db | 435 | QEVE---ELWLGL-----NDLKLQNFWSGSLVPSFTHWPFEPNFRDSDLEDCT 483 | |
| Qy | 124 | MYHOPSPAGIGGPGYMFQWDDRCNMKNFNICKYSDEXKPAVPSREAEG 171 | |
| Db | 484 | IW-----GPEG-----RWNSPCNQSLPSICKKAGQLSQGAAEEDHG 520 | |
| RESULT 13 | | | |
| Q64449 | | | |
| ID | Q64449 | PRELIMINARY; PRT; 1479 AA. | |
| AC | Q64449; | | |
| DT | 01-NOV-1996 | (TrEMBLrel. 01, Created) | |
| DT | 01-NOV-1996 | (TrEMBLrel. 01, Last sequence update) | |
| DT | 01-MAR-2003 | (TrEMBLrel. 23, Last annotation update) | |
| DE | Lectin lambda. | | |
| GN | MRC2. | | |
| OS | Mus musculus (Mouse). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| NCBI_TaxID | 10090; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=96355501; PubMed=8702911; | | |
| RA | Wu K., Yuan J., Laeky L.A.; | | |
| RT | "Characterization of a novel member of the macrophage mannose receptor | | |
| RT | type C lectin family."; | | |
| RL | J. Biol. Chem. 271:21323-21330(1996). | | |
| DR | EMBL; U56734; AAC52729.1; - | | |
| DR | HSSP; P02751; 2FN2. | | |
| DR | MSP; MGI:107818; Mrc2. | | |
| DR | InterPro; IPR002353; AntifreezeII. | | |
| DR | InterPro; IPR001128; Cytochrome P450. | | |
| DR | InterPro; IPR000562; FN Type_II. | | |
| DR | InterPro; IPR001304; Lectin_C. | | |
| DR | InterPro; IPR000566; Lipoclin_CytfABP. | | |
| DR | InterPro; IPR000772; Ricin_B_lectin. | | |
| DR | Pfam; PF00040; fn2; 1. | | |
| DR | Pfam; PF00059; lectin.c; 8. | | |
| DR | PRINTS; PR00356; ANTIFREEZEII. | | |

Search completed: December 22, 2003, 16:13:30
Job time : 25.2017 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:05:39 ; Search time 23.7687 Seconds
(without alignments)
2214.797 Million cell updates/sec

Title: US-09-887-855-2_COPY_24_227

Perfect score: 1106

Sequence: 1 GRLLSGQPVCRGGTQRPCYK.....EEDAKTFKESREAAALNLAY 204

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 1106 | 100.0 | 374 | 4 Q8TAY8 | Q8TAY8 homo sapien |
| 2 | 1106 | 100.0 | 374 | 4 Q96NF3 | Q96NF3 homo sapien |
| 3 | 1102 | 99.6 | 374 | 4 Q96NC5 | Q96NC5 homo sapien |
| 4 | 945.5 | 85.5 | 374 | 11 Q92209 | Q92209 cricetus |
| 5 | 803.5 | 72.6 | 211 | 11 Q8C351 | Q8C351 mus musculus |
| 6 | 580 | 52.4 | 246 | 11 Q8BM17 | Q8BM17 mus musculus |
| 7 | 580 | 52.4 | 292 | 11 Q8BVU2 | Q8BVU2 mus musculus |
| 8 | 185 | 16.7 | 1290 | 13 Q9W6E1 | Q9W6E1 gallus gall |
| 9 | 178.5 | 16.1 | 1456 | 11 Q61830 | Q61830 mus musculus |
| 10 | 177.5 | 16.0 | 1348 | 5 Q25199 | Q25199 hydra atten |
| 11 | 176.5 | 16.0 | 1479 | 4 Q9Y5P9 | Q9Y5P9 homo sapien |
| 12 | 176.5 | 16.0 | 1479 | 4 Q9UBG0 | Q9UBG0 homo sapien |
| 13 | 170.5 | 15.4 | 1479 | 11 Q64449 | Q64449 mus musculus |
| 14 | 169 | 15.3 | 217 | 11 Q8C4F8 | Q8C4F8 mus musculus |
| 15 | 165.5 | 15.0 | 742 | 11 Q8K4Q8 | Q8K4Q8 mus musculus |
| 16 | 165.5 | 15.0 | 742 | 11 Q8C979 | Q8C979 mus musculus |

| | | | | | |
|----|-------|------|------|-----------|---------------------|
| 17 | 161.5 | 14.6 | 134 | 5 Q9XYX3 | Q9XYX3 hydra magni |
| 18 | 159.5 | 14.4 | 742 | 11 Q8VIF6 | Q8VIF6 mus musculus |
| 19 | 158.5 | 14.3 | 719 | 6 Q62623 | Q62623 bos taurus |
| 20 | 158 | 14.3 | 142 | 11 Q8CJ86 | Q8CJ86 mus musculus |
| 21 | 158 | 14.3 | 142 | 11 Q8BHK7 | Q8BHK7 mus musculus |
| 22 | 158 | 14.3 | 295 | 11 Q912W4 | Q912W4 mus musculus |
| 23 | 158 | 14.3 | 311 | 11 Q9D8V4 | Q9D8V4 mus musculus |
| 24 | 158 | 14.3 | 325 | 11 Q912X0 | Q912X0 mus musculus |
| 25 | 157 | 14.2 | 158 | 13 Q90W17 | Q90W17 bungarus fa |
| 26 | 156 | 14.1 | 323 | 11 Q8CJ91 | Q8CJ91 mus musculus |
| 27 | 156 | 14.1 | 339 | 6 Q95244 | Q95244 sus scrofa |
| 28 | 155 | 14.0 | 293 | 11 Q8BG20 | Q8BG20 mus musculus |
| 29 | 155 | 14.0 | 323 | 11 Q8CJ94 | Q8CJ94 mus musculus |
| 30 | 155 | 14.0 | 323 | 11 Q8CJ93 | Q8CJ93 mus musculus |
| 31 | 155 | 14.0 | 323 | 11 Q8CJ88 | Q8CJ88 mus musculus |
| 32 | 155 | 14.0 | 1152 | 13 Q90WM2 | Q90WM2 xenopus lae |
| 33 | 154 | 13.9 | 322 | 11 Q8CJ89 | Q8CJ89 mus musculus |
| 34 | 154 | 13.9 | 323 | 11 Q8CJ92 | Q8CJ92 mus musculus |
| 35 | 152.5 | 13.8 | 485 | 6 Q95LG3 | Q95LG3 odocoileus |
| 36 | 152 | 13.7 | 158 | 13 Q90W16 | Q90W16 bungarus mu |
| 37 | 151.5 | 13.7 | 652 | 4 Q81XK1 | Q81XK1 homo sapien |
| 38 | 150.5 | 13.6 | 399 | 6 Q8HY12 | Q8HY12 hylobates l |
| 39 | 150 | 13.6 | 197 | 6 Q28008 | Q28008 bos taurus |
| 40 | 149.5 | 13.5 | 158 | 11 Q8JZX6 | Q8JZX6 mus musculus |
| 41 | 149.5 | 13.5 | 459 | 5 Q22136 | Q22136 caenorhabdi |
| 42 | 149 | 13.5 | 158 | 13 Q90W18 | Q90W18 bungarus fa |
| 43 | 148.5 | 13.4 | 163 | 13 Q8AXR8 | Q8AXR8 anguilla ja |
| 44 | 148.5 | 13.4 | 195 | 5 Q27340 | Q27340 megabalanus |
| 45 | 148.5 | 13.4 | 330 | 4 Q9NT67 | Q9NT67 homo sapien |

ALIGNMENTS

RESULT 1

Q8TAY8 ID Q8TAY8 PRELIMINARY; PRT; 374 AA.

AC Q8TAY8;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Similar to unnamed protein product.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RE EMBL; BC025407; AAH25407.1;
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00041; C TYPE LECTIN 2; 1.
SQ SEQUENCE 374 AA; 42312 MW; EC214E6BC9E578D9 CRC64;

Query Match 100.0%; Score 1106; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 2.8e-98;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| Qy | 1 | GRLLSGQPVCRGGTQRPCYKVIYFHDTSRLNFEAKACRRDGGQLVSTSEDEQKLE | 60 |
| Db | 24 | GRLLSGQPVCRGGTQRPCYKVIYFHDTSRLNFEAKACRRDGGQLVSTSEDEQKLE | 83 |
| Qy | 61 | KFIENLLPSDGFWMIGLRRREEKQSNSTACQDLYAWTDGSIQPRNNYVDEPSCGSEVCV | 120 |
| Db | 84 | KFIENLLPSDGFWMIGLRRREEKQSNSTACQDLYAWTDGSIQPRNNYVDEPSCGSEVCV | 143 |
| Qy | 121 | VMHQSPAPAGIGGPPYFQWDDRCNNKNNPKYISDEKPAVPSREAGSETELTTPVLP | 180 |
| Db | 144 | VMHQSPAPAGIGGPPYFQWDDRCNNKNNPKYISDEKPAVPSREAGSETELTTPVLP | 203 |

| | |
|---|--|
| RA | Tashiro H., Yamazaki M., Watanabe K., Kunagai A., Itakura S., |
| RA | Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R., |
| RA | Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., |
| RA | Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., |
| RA | Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M., |
| RA | Murakami K., Kanehori K., Takahashi-Fujii A., Oshina A., Sugiyama A., |
| RA | Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuhara Y., Nagai K., |
| RA | Isozaki T.; |
| RT | "NEDO human cDNA sequencing project."; |
| RL | Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases. |
| DR | EMBL; AK055654; BAB70978.1; - |
| DR | InterPro; IP001304; Lectin_C. |
| DR | Pfam; PF00059; Lectin_c; 1. |
| DR | SMART; SM00034; CLECT; 1. |
| DR | PROSITE; PS00041; C_TYPE_LECTIN_2; 1. |
| KW | Hypothetical protein. |
| SQ | SEQUENCE 374 AA; 42310 MW; CBF74B67E23BA16 CRC64; |
| | |
| Query Match | 99.6%; Score 1102; DB 4; Length 374; |
| Best Local Similarity | 99.5%; Pred. No. 6.8e-98; |
| Matches 203; Conservative | 0; Mismatches 1; Indels 0; Gaps 0 |
| | |
| Qy 1 | GRLLSGQPVCRGGTORPCYKVIYPHDTSRRLNFEEAKACRRDGGOLVSIESDEOKLIE 60 |
| Dd | 24 GRLLSGQPVCRGGTORPCYKVIYPHDTSRRLNFEEAKACRRDGGOLVSIESDEOKLIE 83 |
| | |
| Qy 61 | KFIENLLPSDGFVIGLRREKQSNSTACQDIYAWTDGSIQFRWYVDSPSCGSEVCV 120 |
| Dd | 84 KFIENLLPSDGFVIGLRREKQSNSTACQDIYAWTDGSIQFRWYVDSPSCGSEVCV 143 |
| | |
| Qy 121 | VMYHQPSAPAGIGGPYMFQWDDRCNMKNFNICKYSDEKPAVPSPREAGEETELTTPVLP 180 |
| Dd | 144 VMYHQPSAPAGIGGPYMFQWDDRCNMKNFNICKYSDEKPAVPSPREAGEETELTTPVLP 203 |
| | |
| Qy 181 | BETOEDDAKTFTKSREAAALNAY 204 |
| Dd | 204 BETOEEDTKTFTKSREAAALNAY 227 |
| | |
| RESULT 4 | |
| Q9Z209 | PRELIMINARY; PRT; 374 AA. |
| ID Q9Z209 | |
| AC Q9Z209; | |
| DT 01-MAY-1999 | (TrEMBLrel. 10, Created) |
| DT 01-MAY-1999 | (TrEMBLrel. 10, Last sequence update) |
| DT 01-MAR-2003 | (TrEMBLrel. 23, Last annotation update) |
| DE Laylin. | |
| OS Cricetus griseus (Chinese hamster). | |
| OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; | |
| OC Cricetulus. | |
| OX NCBI_TaxID=10029; | |
| RN [1] | |
| RP SEQUENCE FROM N.A. | |
| RA Borowsky M.L., Hynes R.O.; | |
| RT "Laylin, a novel talin-binding transmembrane protein homologous with | |
| RT C-type lectins, is localized in membrane ruffles."; | |
| RL J. Cell Biol. 143:0-0(1998). | |
| DR EMBL; AF093673; AAC68695.1; - | |
| DR HSPF; P22897; 1EGG. | |
| DR InterPro; IP001304; Lectin_C. | |
| DR Pfam; PF00059; lectin_c; 1. | |
| DR SMART; SM00034; CLECT; 1. | |
| DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1. | |
| SQ SEQUENCE 374 AA; 42435 MW; Z98A8BA24PB04E1C CRC64; | |
| | |
| Query Match | 85.5%; Score 945.5; DB 11; Length 374; |
| Best Local Similarity | 85.9%; Pred. No. 8.2e-83; |
| Matches 176; Conservative | 9; Mismatches 19; Indels 1; Gaps 1 |
| | |
| Qy 1 | GRLLSGQPVCRGGTORPCYKVIYPHDTSRRLNFEEAKACRRDGGOLVSIESDEOKLIE 60 |
| Dd | 24 GRLLSGQPVCRGGTORPCYKVIYPHDTSRRLNFEEAKACRRDGGOLVSIESDEOKLIE 83 |


```
QY 61 KFIENLLPSDGFWMIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVDEPSCGSEVCV 120
Db 84 KFIENLLASDGFWMIGLRLLEVKQVNTACQDLYAWTDGSIQFRNMYVDEPSCGSEVCV 143
QY 121 VMYHQPAPAGIGGYPYFQWDDRCNMKNFNICKYSDKPA-VPSREABGEETELTPVL 179
Db 144 VMYHQPAPAGIGGYPYFQWDDRCNMKNFNICKYADEKPTTIPRPGGEATEPPTPVL 203
QY 180 PEETQEDDAKTKFKESREAAINLAY 204
Db 204 PEETQEDTKETFKESREAAINLAY 228

RESULT 5
Q8C351 PRELIMINARY; PRT; 211 AA.
ID AC Q8C351;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE LAYILIN homolog (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).";
DR EMBL; AK086930; BAC39765.1; -.
FT NON_TER 211
SQ SEQUENCE 211 AA; 23697 MW; AD9870B5957DD5AE CRC64;

Query Match 72.6%; Score 803.5; DB 11; Length 211;
Best Local Similarity 79.8%; Pred. No. 2e-69;
Matches 150; Conservative 9; Mismatches 20; Indels 9; Gaps 2;

QY 1 GRLLS-----GQVCRGGTQPCYKVIYFHDTSRRRLNFEAKACRRDGGQLVSIRES 52
Db 24 GRLLSASDLDPGGQVCRGGTQPCYKVIYFHDFAFQRLNFEAKETCMEDGGQLVSIET 83
QY 53 EDEOKLIEFIENLLPSDGFWMIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVDEP 112
Db 84 EDEORLIEFIENLLASDGFWMIGLRLLEKQSNNTACQDLYAWTDGSIQFRNMYVDEP 143
QY 113 SCGSEVCVVMYHQPAPAGIGGYPYFQWDDRCNMKNFNICKYSDKPA-VPSREAGEE 171
Db 144 SCGSEVCVVMYHQPAPAGIGGYPYFQWDDRCNMKNFNICKYHDDKPTTIPRPGGEA 203
QY 172 TELTTPVL 179
Db 204 TEPATPLL 211

RESULT 6
Q8BM17 PRELIMINARY; PRT; 246 AA.
ID AC Q8BM17;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE C-type lectin protein MT75 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Porelimb;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).";
DR EMBL; AK031063; BAC27234.1; -.
SQ SEQUENCE 246 AA; 27394 MW; E872660C58267752 CRC64;

Query Match 52.4%; Score 580; DB 11; Length 246;
Best Local Similarity 57.9%; Pred. No. 8.1e-48;
Matches 110; Conservative 26; Mismatches 40; Indels 14; Gaps 5;

QY 2 RLLSGQPVCRGGTQPCYKVIYFHDTSRRRLNFEAKACRRDGGQLVSIRESDEOKLIEK 61
Db 23 RVVSGQKVCFAADVKKPCYKVIYFHELSRVFQEARLACESGGVLLSLENEAEQKLIES 82
QY 62 FIENLLP-----SDGDFWIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVDEPSCGS 116
Db 83 MLQNLTKPGTGISDGFWMIGLRLSFGDQT-SGACPDLYQWSDGSSQFRNMYTDEPSCGS 141
QY 117 EVCVVMYHQPAPAGIGGYPYFQWDDRCNMKNFNICKYSDKPA-VPSREAGEEETELT 175
Db 142 EKCVMYHQPAPAGIGGYPYFQWDDRCNMKNFNICKYKPEIHTPEA-----EKPILT 196
QY 176 TPVLPEETQE 185
Db 197 NQ--PEETHE 204

RESULT 7
Q8BVU2 PRELIMINARY; PRT; 292 AA.
ID AC Q8BVU2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE C-type lectin protein MT75 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).";
DR EMBL; AK076523; BAC36378.1; -.
SQ SEQUENCE 292 AA; 32502 MW; 73B631C0714D54E2 CRC64;

Query Match 52.4%; Score 580; DB 11; Length 292;
Best Local Similarity 57.9%; Pred. No. 1e-47;
Matches 110; Conservative 26; Mismatches 40; Indels 14; Gaps 5;

QY 2 RLLSGQPVCRGGTQPCYKVIYFHDTSRRRLNFEAKACRRDGGQLVSIRESDEOKLIEK 61
Db 23 RVVSGQKVCFAADVKKPCYKVIYFHELSRVFQEARLACESGGVLLSLENEAEQKLIES 82
QY 62 FIENLLP-----SDGDFWIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVDEPSCGS 116
Db 83 MLQNLTKPGTGISDGFWMIGLRLSFGDQT-SGACPDLYQWSDGSSQFRNMYTDEPSCGS 141
QY 117 EVCVVMYHQPAPAGIGGYPYFQWDDRCNMKNFNICKYSDKPA-VPSREAGEEETELT 175
Db 142 EKCVMYHQPAPAGIGGYPYFQWDDRCNMKNFNICKYKPEIHTPEA-----EKPILT 196
QY 176 TPVLPEETQE 185
```

Db 197 NQ--PEETHE 204

|||||

RESULT 8

Q9W6E1 PRELIMINARY; PRT; 1290 AA.

ID Q9W6E1

AC Q9W6E1

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Neurocan core protein.

OS Gallus gallus (Chicken)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archoosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20309833; PubMed=10851024;

RA Li H., Leung T.C., Hoffman S., Balsamo J., Lilien J.;

RT "Coordinate Regulation of Cadherin and Integrin Function by the

RL Chondroitin Sulfate Proteoglycan Neurocan.";

RL J. Cell Biol. 149:1275-1288 (2000).

DR EMBL; AF116856; AAD24546.2; -.

DR HSP; P08709; 1BF9.

DR InterPro; IPR000152; Asx_hydroxyl.

DR InterPro; IPR000742; EGF 2.

DR InterPro; IPR001881; EGF_Ca.

DR InterPro; IPR001438; EGF-II.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR003599; Ig.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR001304; Lectin_C.

DR InterPro; IPR000538; Link.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00008; EGF; 2.

DR Pfam; PF00047; Ig; 1.

DR Pfam; PF00059; lectin_c; 1.

DR Pfam; PF00084; sushi; 1.

DR Pfam; PF00193; Xlink; 2.

DR PRINTS; PR00010; EGFBLD.

DR PRINTS; PR0265; LINKMODULE.

DR PRODOM; PD000918; Link; 2.

DR SMART; SM00032; CCP; 1.

DR SMART; SM00034; CLECT; 1.

DR SMART; SM00179; EGF_CA; 1.

DR SMART; SM00409; IG_1.

DR SMART; SM00445; LINK; 2.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.

DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.

DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.

DR PROSITE; PS00022; EGF_1; 2.

DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS01187; EGF_CA; 1.

DR PROSITE; PS00835; IG_LIKE; 1.

DR PROSITE; PS01241; LINK; 2.

KW EGF-like domain.

SQ SEQUENCE 1290 AA; 138877 MW; 182BD86D0E40BE78 CRC64;

Query Match 16.7%; Score 185; DB 13; Length 1290;

Best Local Similarity 34.7%; Pred. No. 6.8e-09;

Matches 50; Conservative 18; Mismatches 42; Indels 34; Gaps 8;

QY 15 QRCYKVIYFHDTSRLNPFEEAKACRRDGGQLVSISEDEQKLIBKFTENLPSDGDFFW 74

Db 1064 QHCYR--YF---SRRSWEDAECDRCRRAGHTLSHSQEEHGFINSF-----GHENTW 1112

QY 75 IGLRRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCVMY-HOPSAPA 130

Db 1113 IGLNDRIVEQD-----FQWTDNTGLQYENWRENQPNFFAGGDCVLSVSH----- 1159

197 NQ--PEETHE 204

|||||

RESULT 9

Q61830 PRELIMINARY; PRT; 1456 AA.

ID Q61830

AC Q61830

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Macrophage mannose receptor precursor.

GN MRC1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=C57BL/6;

RX MEDLINE=93043353; PubMed=1421407;

RA Harris N., Rits M., Chang G., Ezekowitz R.B.;

RT "Characterization of the murine macrophage mannose receptor.";

RL Blood 80:2363-2373 (1992).

RL [2]

RP SEQUENCE FROM N.A.

RX STRAIN=C57BL/6;

RA Super M.;

RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z11974; CAA78028.1; -.

DR HSP; P22897; 1EGG.

DR MGD; MGI:971142; Mrcl.

DR InterPro; IPR002353; AntifreezeII.

DR InterPro; IPR000562; FN_Type_II.

DR InterPro; IPR001304; Lectin_C.

DR InterPro; IPR000772; Ricin_B_lectin.

DR Pfam; PF00040; fn2; 1.

DR Pfam; PF00059; lectin_c; 8.

DR Pfam; PF00652; Ricin_B_lectin; 2.

DR PRINTS; PR00356; ANTIFREEZEII.

DR PRINTS; PR00013; FNYPEII.

DR PRODOM; PD000995; FN_Type_II; 1.

DR SMART; SM00034; CLECT; 8.

DR SMART; SM00059; FN2; 1.

DR SMART; SM00458; RICIN; 1.

DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 6.

DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 8.

DR PROSITE; PS00023; FIBRONECTIN_2; 1.

DR PROSITE; PS02031; RICIN_B_LLECTIN; 1.

KW Receptor; Signal.

FT CHAIN 1 18 POTENTIAL.

FT CHAIN 19 1456 MACROPHAGE MANNOSE RECEPTOR.

SQ SEQUENCE 1456 AA; 165065 MW; 4EBD3F1B8619A594 CRC64;

Query Match 16.1%; Score 178.5; DB 11; Length 1456;

Best Local Similarity 25.9%; Pred. No. 3.3e-08;

Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;

QY 19 YKVIYFHDTSRLNPFEEAKACRRDGGQLVSISEDEQKLIBKFTENLPSDGDFFW 78

Db 807 YKDYQYFYFSKEKETMDNARFCCKNFGLATIKSESEKFLWKYI-NKNGGQSPYFGML 865

QY 79 RREBKQSNSTACQDLYAWTDGSIQFRNMYVDEPSGCS--EVCVMYHQSPAPAGIGPY 136

Db 866 ISWDKK-----FIWMDGSKVDVFNATGEPFNANDENCVTWY-----TNSGF----- 908

QY 137 MFQWDDRCMKKNFICK---YSDEKPAVPSREAGEETELTPVLPEETQE----- 185

Db 909 ---WINDINCQYNNFCQRHNSINATAMP-----TTPTPGCKEGWHLYKNK 954

QY 186 -----EDAKTFKESREALNL 202

197 NQ--PEETHE 204

|||||

RESULT 8

Q9W6E1 PRELIMINARY; PRT; 1290 AA.

ID Q9W6E1

AC Q9W6E1

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Neurocan core protein.

OS Gallus gallus (Chicken)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archoosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20309833; PubMed=10851024;

RA Li H., Leung T.C., Hoffman S., Balsamo J., Lilien J.;

RT "Coordinate Regulation of Cadherin and Integrin Function by the

RL Chondroitin Sulfate Proteoglycan Neurocan.";

RL J. Cell Biol. 149:1275-1288 (2000).

DR EMBL; AF116856; AAD24546.2; -.

DR HSP; P08709; 1BF9.

DR InterPro; IPR000152; Asx_hydroxyl.

DR InterPro; IPR000742; EGF 2.

DR InterPro; IPR001881; EGF_Ca.

DR InterPro; IPR001438; EGF-II.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR003599; Ig.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR001304; Lectin_C.

DR InterPro; IPR000538; Link.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00008; EGF; 2.

DR Pfam; PF00047; Ig; 1.

DR Pfam; PF00059; lectin_c; 1.

DR Pfam; PF00084; sushi; 1.

DR Pfam; PF00193; Xlink; 2.

DR PRINTS; PR00010; EGFBLD.

DR PRINTS; PR0265; LINKMODULE.

DR PRODOM; PD000918; Link; 2.

DR SMART; SM00032; CCP; 1.

DR SMART; SM00034; CLECT; 1.

DR SMART; SM00179; EGF_CA; 1.

DR SMART; SM00409; IG_1.

DR SMART; SM00445; LINK; 2.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.

DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.

DR PROSITE; PS

Db 955 CFKIFGFANEKKSGWQARQACKGL 979

RESULT 10

ID Q25199 PRELIMINARY; PRT; 1348 AA.

AC Q25199;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Tyrosine kinase receptor.

OS Hydra attenuata (Hydra) (Hydra vulgaris).

OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Anthomedusae;

OC Hydrozoa; Hydra.

OX NCBI_TaxID=6087;

FN [1]

RN SEQUENCE FROM N.A.

RP STRAIN=IrVine;

RC MEDLINE=20209407; PubMed=10744720;

RA Reidling J.C., Miller M.A., Steele R.E.;

RT "Sweet Tooth, a Novel Receptor Protein-Tyrosine Kinase with C-type

RT Lactin-like Extracellular Domains.";

RL J. Biol. Chem. 275:10323-10330(2000).

DR EMBL; L22612; AAA29218.2; -.

DR HSP; P11362; IFGK.

DR InterPro; IPR001304; Lactin_C.

DR InterPro; IPR003990; Pancreatins.ac.

DR InterPro; IPR00719; Prot.kinase.

DR InterPro; IPR001245; Tyr_kinase.

DR Pfam; PF00059; lectin_C; 4.

DR Pfam; PF00069; pkinase; 1.

DR PRINTS; PD00001; PNCREATINSAP.

DR ProDom; PD000001; Prot.kinase; 2.

DR SMART; SM00034; CLECT; 4.

DR SMART; SM00219; TyKc; 1.

DR PROSITE; PS00615; C-TYPE LECTIN 1; 2.

DR PROSITE; PS00041; C-TYPE LECTIN_2; 4.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.

DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.

KW ATP-binding; Kinase; Receptor; Transferase; Tyrosine-protein kinase.

SQ SEQUENCE 1348 AA; 156916 MW; 77D2122093227FEP CRC64;

Query Match 16.0%; Score 177.5; DB 5; Length 1348;

Best Local Similarity 28.0%; Pred. No. 3.8e-08;

Matches 47; Conservative 32; Mismatches 46; Indels 43; Gaps 8;

QY 4 LSGQPV-C-RGGTQRPC-----YKVIYFHDTSRLNFEAEKACRRDGGQLVSISEDE 55

Db 412 LSHRFCKVKRATNEYCAEGWTSYRIYCFIYSIEFDWFKSFSCQNGNLLSIENQEE 471

QY 56 QKLIKIEFIENLLPSDGD-FWIGLRR-----REEKQSNSTACQDLYAWTDGSIQFRNMY 108

Db 472 ---NRFIENDLIKNDKYWIGLNKIWNLYKKNR-----FWSQNTYTFQFNWI 518

QY 109 VDEP--SCGSEVCVVMYHQSAPAGIGPYFMQNDRCNMKNFICK 154

Db 519 TNOPDNNNGIESCVEMYN-----GMSDKCKVLNGFICK 553

RESULT 11

QY5P9 PRELIMINARY; PRT; 1479 AA.

ID QY5P9;

AC QY5P9;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Endocytic receptor Endo180.

GN ENDO180.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa;

OC Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN SEQUENCE FROM N.A.

RP MEDLINE=20148849; PubMed=10683150;

RA Sheikh H., Yarwood H., Ashworth A., Isacke C.;

RT "Endo180, an endocytic recycling glycoprotein related to the

RT macrophage mannose receptor is expressed on fibroblasts, endothelial

RL cells and macrophages and functions as a lectin receptor.";

DR EMBL; AF134838; AAD30280.1; -.

DR HSP; P02751; 2FN2

DR InterPro; IPR000562; FN_Type_II.

DR InterPro; IPR001304; Lactin_C.

DR InterPro; IPR000566; Lipocln.cytPABP.

DR InterPro; IPR000772; Ricin_B_lectin.

DR Pfam; PF00040; fn2; 1.

DR Pfam; PF00059; lectin_C; 8.

DR PRINTS; PR00013; FNTYPEII.

DR ProDom; PD000995; FN_Type_II; 1.

DR SMART; SM00034; CLECT; 8.

DR SMART; SM00059; FN2; 1.

DR SMART; SM00458; RICIN; 1.

DR PROSITE; PS00615; C-TYPE LECTIN 1; 3.

DR PROSITE; PS00041; C-TYPE LECTIN_2; 8.

DR PROSITE; PS00023; FIBRONECTIN 2; 1.

DR PROSITE; PS00213; LIPOCALIN; 1.

DR PROSITE; PS00231; RICIN_B_LECTIN; 1.

KW Receptor.

SQ SEQUENCE 1479 AA; 166669 MW; 9F4BAF355F036FCE CRC64;

Query Match 16.0%; Score 176.5; DB 4; Length 1479;

Best Local Similarity 30.4%; Pred. No. 5.3e-08;

Matches 51; Conservative 25; Mismatches 57; Indels 35; Gaps 7;

QY 5 SQQPVCGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSISEDEOKLIEKPIE 64

Db 385 SWQPF-----QGHCYRL-----QAERKSWQESKACLRGGDLVSIHSMAELEFITKQIK 434

QY 65 NLLPSDGDGFWIGLRRREEKQSNSTACQDLYAWTDGSIQFRNMYVDPS---CGSEVCV 121

Db 435 QVEV---ELWIGL-----NDLKQMFWSQSLVSTHWHPPFNPRDSLEDCT 483

QY 122 MYHQPSAPAGIGPYFMQNDRCNMKNFICKYSDKPAVPSREAG 169

Db 484 IW---GPEG-----RWNSPCNQSLPSICKKAGQLSQGAEDHG 520

RESULT 12

QYUBGO PRELIMINARY; PRT; 1479 AA.

ID QYUBGO;

AC QYUBGO;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Urokinase receptor-associated protein UPARAP.

GN KIAA0709.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

FN [1]

RN SEQUENCE FROM N.A.

RP Behrendt N., Jensen O.N., Engelholm L.H., Mortz E., Mann M., Dano K.;

RT "A urokinase receptor-associated protein with specific collagen-

RT binding properties.";

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

RN SEQUENCE FROM N.A.

RP TISSUE=Brain;

RC MEDLINE=98403880; PubMed=9734811;

RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,

RA Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. X.

RT The complete sequences of 100 new cDNA clones from brain which can


```

Q8K4Q8      Q8K4Q8      PRELIMINARY;      PRT;      742 AA.
AC      Q8K4Q8;
DT      01-OCT-2002 (TRENBLrel. 22, Created)
DT      01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT      01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE      Collectin placenta 1.
GN      COLEC12 OR CL-PI.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Ohtani K., Suzuki Y., Eda S., Kawai T., Kase T., Keshi H., Sakai Y.,
RA      Fukuh A., Sakamoto T., Itabe H., Suzutani T., Ogasawara M.,
RA      Yoshida I., Wakamiya N.;
RT      "cDNA cloning of mouse CL-PI gene.";
RL      Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AB078434; BAC05523.1; -.
DR      MGD; MGI:2152907; Colecl2.
DR      InterPro; IPR002353; AntifreezeII.
DR      InterPro; IPR000087; Collagen.
DR      InterPro; IPR001304; Lectin_C.
DR      Pfam; PF01391; Collagen; 3.
DR      Pfam; PF00059; lectin_c; 1.
DR      PRINTS; PR00356; ANTIFREEZEII.
DR      SMART; SM00034; CLECT; 1.
DR      PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR      PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
SQ      SEQUENCE 742 AA; 81304 MW; 1537C490B5911C45 CRC64;

Query Match      15.0%; Score 165.5; DB 11; Length 742;
Best Local Similarity 33.8%; Pred. No. 2.6e-07;
Matches 50; Conservative 17; Mismatches 54; Indels 27; Gaps 8;

Qy      18 CYKVIYFHTSRLNFEAKACRRDGGQLVSEDEQKLIKFIENLLPSDGDGFWIGL 77
Db      ||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
618 CY---YF--SLEKEIFEDAKLFCEDKSHLVFINSREEQWIKKH---TVGRESHWIGL 668

Qy      78 RRREEKQSNSTACDLYAWTDGSIQPRNVTYVDEP-SCGSEVCVVMYHQPSAPAGIGGPY 136
Db      ||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
669 TDSQESE-----WKWLDGSPVDYKNWKGQPDNWSG-----HGPGEDCA-GLIY 713

Qy      137 MFQWDDRCNMKNFNICKYSDKPAVPS 164
Db      ||| : : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
714 AGQWDFQCDIINNICE--KEREAVPS 739

```

Search completed: December 22, 2003, 16:13:30
 Job time : 23.9687 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:05:09 ; Search time 6.56161 Seconds
(without alignment)
1476.391 Million cell updates/sec

Title: US-09-887-855-5

Perfect score: 1115

Sequence: 1 ATGRLSGQPVCRCGTORPC.....BEDAKKTFKESREALNLAY 206

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 574 | 51.4 | 273 | 1 | CHOD_MOUSE |
| 2 | 561.5 | 50.4 | 273 | 1 | Q9h9p2 mus musculus |
| 3 | 182 | 16.3 | 1456 | 1 | MANR_HUMAN |
| 4 | 177 | 15.9 | 1268 | 1 | PCGN_MOUSE |
| 5 | 176 | 15.8 | 1321 | 1 | PCGN_HUMAN |
| 6 | 174.5 | 15.7 | 3381 | 1 | PCGV_BOVIN |
| 7 | 174 | 15.6 | 1257 | 1 | PCGN_RAT |
| 8 | 174 | 15.6 | 2738 | 1 | PCGV_MOUSE |
| 9 | 174 | 15.6 | 3358 | 1 | PCGV_HUMAN |
| 10 | 174 | 15.6 | 3396 | 1 | PCGV_CHICK |
| 11 | 171 | 15.3 | 3562 | 1 | C993_RAT |
| 12 | 165 | 14.8 | 643 | 1 | PCGA_BOVIN |
| 13 | 158.5 | 14.2 | 2364 | 1 | PCGA_HUMAN |
| 14 | 158.5 | 14.2 | 2415 | 1 | PCGA_CANFA |
| 15 | 155.5 | 13.9 | 2333 | 1 | PCGA_MOUSE |
| 16 | 154.5 | 13.9 | 612 | 1 | C993_MOUSE |
| 17 | 153.5 | 13.8 | 644 | 1 | PCGA_RAT |
| 18 | 153.5 | 13.8 | 2124 | 1 | PCGA_BOVIN |
| 19 | 152 | 13.6 | 912 | 1 | C993_HUMAN |
| 20 | 151.5 | 13.6 | 652 | 1 | PCF2_HUMAN |
| 21 | 151 | 13.5 | 197 | 1 | CLF1_HUMAN |
| 22 | 149 | 13.4 | 321 | 1 | PCGA_MOUSE |
| 23 | 148.5 | 13.3 | 2132 | 1 | PCGB_MOUSE |
| 24 | 147 | 13.2 | 883 | 1 | PCGB_MOUSE |
| 25 | 146 | 13.1 | 2109 | 1 | PCGA_CHICK |
| 26 | 145 | 13.0 | 158 | 1 | LECG_TRIST |
| 27 | 145 | 13.0 | 883 | 1 | PCGB_RAT |
| 28 | 144.5 | 13.0 | 173 | 1 | LEC2_MEGRO |
| 29 | 144.5 | 13.0 | 372 | 1 | LEMI_RAT |
| 30 | 143.5 | 12.9 | 372 | 1 | LEMI_MOUSE |
| 31 | 141.5 | 12.7 | 331 | 1 | FCE2_MOUSE |
| 32 | 141.5 | 12.7 | 549 | 1 | LEMI_RAT |
| 33 | 140.5 | 12.6 | 162 | 1 | LEC3_MEGRO |
| | | | | | P07439 megabalanus |

ALIGNMENTS

RESULT 1

CHOD_MOUSE
ID CHOD_MOUSE STANDARD; PRT: 273 AA.
AC Q9CXM0; Q8VI31;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chondrolectin precursor (Transmembrane protein MT75).
GN CHODL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Weng L., Smith P., Hubner R., Wouters J., Merregaert J.;
RT "Mt75, a low expressed c-type lectin gene involving in
chondrogenesis.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Matzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

EMBL; AF311699; AAU50354.1; -

P98131 bos taurus
P07714 homo sapien
P05047 sarcophaga
P23806 trimeresuru
P43025 mus musculus
Q95198 macaca mula
Q28768 papio hamad
P23132 bos taurus
P43137 mus musculus
P83300 anser anser
Q95235 pongo pygma
P98107 bos taurus


```

DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE LECTIN 1; FALSE_NEG.
DR PROSITE: PS50041; C-TYPE LECTIN 2; 1.
KW Lectin; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 273 CHONDROLECTIN.
FT DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 217 237 POTENTIAL.
FT DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 179 C-TYPE LECTIN.
FT CARBOHYD 86 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 273 AA; 30431 MW; F4890AAB572A311 CRC64;

Query Match 50.4%; Score 561.5; DB 1; Length 273;
Best Local Similarity 60.1%; Pred. No. 5.7e-43;
Matches 101; Conservative 25; Mismatches 35; Indels 7; Gaps 3;

QY 4 RLLSGOPVCRGGTQPCYKVIYFHDTSRLNFEAKACRRDGGQLVSISEDEQKLEK 63
DB 23 RVVSGQKVCFADPKHPCKYKWAYFHEUSSRVFQEARLACSEGGVLLSLENAEQKLES 82

QY 64 FIENLLP-----SDGFWIGLRRREKQSNSTACQDLYAMTDSISOFRNWYVDEPSCGS 118
DB 83 MLQNLTKPGTGISDGDWFGLWRNGDQT-SGACPDLYQWSDGNSQYRNWYTDPSGCS 141

QY 119 EVCVMYHQPSAPAGTGGPYFMQWDDRCNNKKNFKICKYSDE-KPAVP 165
DB 142 EKCVMYHQPTANPLGLGPLYQWDDRCNNKKNYICKYEPINPTAP 189

RESULT 3
MANR_HUMAN STANDARD; PRT; 1456 AA.
ID AC P22897;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Macrophage mannose receptor precursor (MMR) (CD206 antigen).
GN MRC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=90324192; PubMed=2373685;
RA Taylor M.E., Conary J.T., Lennartz M.R., Stahl P.D., Drickamer K.;
RT "Primary structure of the mannose receptor contains multiple motifs
RT resembling carbohydrate-recognition domains.";
RL J. Biol. Chem. 265:12156-12162(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93052405; PubMed=1294119;
RA Kim S.J., Ruiz N., Bezouska K., Drickamer K.;
RT "Organization of the gene encoding the human macrophage mannose
RT receptor (MRC1).";
RL Genomics 14:721-727(1992).
RN [3]
RP STUDIES ON THE BINDING OF INDIVIDUAL LECTIN DOMAINS.
RX MEDLINE=92112893; PubMed=1730714;
RA Taylor M.E., Bezouska K., Drickamer K.;
RT "Contribution to ligand binding by multiple carbohydrate-recognition
RT domains in the macrophage mannose receptor.";
RL J. Biol. Chem. 267:1719-1726(1992).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 642-788.
RX MEDLINE=20347275; PubMed=10779515;
RA Feinberg H., Park-Snyder S., Kolatkar A.R., Heise C.T., Taylor M.E.,
RA Weis W.I.;
RT "Structure of a C-type carbohydrate recognition domain from the
RT macrophage mannose receptor.";

```

```

RL J. Biol. Chem. 275:21539-21548(2000).
CC -I- FUNCTION: MEDIATES THE ENDOCYTOSIS OF GLYCOPROTEINS BY
CC MACROPHAGES, IN SEVERAL RECOGNITION AND UPTAKE PROCESSES.
CC -I- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -I- MISCELLANEOUS: CRDs 1-3 HAVE AT MOST VERY WEAK AFFINITY FOR
CC CARBOHYDRATE. CRD 4 SHOWS THE HIGHEST AFFINITY BINDING AND HAS
CC MULTISPECIFICITY FOR A VARIETY OF MONOSACCHARIDES. AT LEAST 3 CRDS
CC ENDOCYTOSIS OF MULTIVALENT GLYCOCONJUGATES.
CC -I- SIMILARITY: Contains 8 C-type lectin family domains.
CC -I- DATABASE: NAME=PROW; NOTE=PROW 2:85-89(2001);
CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1644341535_g.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: J05550; AAA59868.1;
DR EMBL: M93221; AAA60389.1;
DR EMBL: M93192; AAA60389.1; JOINED.
DR EMBL: M93193; AAA60389.1; JOINED.
DR EMBL: M93194; AAA60389.1; JOINED.
DR EMBL: M93195; AAA60389.1; JOINED.
DR EMBL: M93196; AAA60389.1; JOINED.
DR EMBL: M93197; AAA60389.1; JOINED.
DR EMBL: M93198; AAA60389.1; JOINED.
DR EMBL: M93199; AAA60389.1; JOINED.
DR EMBL: M93200; AAA60389.1; JOINED.
DR EMBL: M93201; AAA60389.1; JOINED.
DR EMBL: M93202; AAA60389.1; JOINED.
DR EMBL: M93203; AAA60389.1; JOINED.
DR EMBL: M93204; AAA60389.1; JOINED.
DR EMBL: M93205; AAA60389.1; JOINED.
DR EMBL: M93206; AAA60389.1; JOINED.
DR EMBL: M93207; AAA60389.1; JOINED.
DR EMBL: M93208; AAA60389.1; JOINED.
DR EMBL: M93209; AAA60389.1; JOINED.
DR EMBL: M93210; AAA60389.1; JOINED.
DR EMBL: M93211; AAA60389.1; JOINED.
DR EMBL: M93212; AAA60389.1; JOINED.
DR EMBL: M93213; AAA60389.1; JOINED.
DR EMBL: M93214; AAA60389.1; JOINED.
DR EMBL: M93215; AAA60389.1; JOINED.
DR EMBL: M93216; AAA60389.1; JOINED.
DR EMBL: M93217; AAA60389.1; JOINED.
DR EMBL: M93218; AAA60389.1; JOINED.
DR EMBL: M93219; AAA60389.1; JOINED.
DR EMBL: M93220; AAA60389.1; JOINED.
DR FIR: A36563; A36563.
DR PDB: 1EGG; 30-AUG-00.
DR PDB: 1EGI; 30-AUG-00.
DR Genew: HGNC:7228; MRC1.
DR MIM: 153618;
DR GO: GO:0005887; C:integral to plasma membrane; TAS.
DR GO: GO:0005537; F:mannose binding activity; TAS.
DR GO: GO:0004872; F:receptor activity; TAS.
DR GO: GO:0006898; F:receptor mediated endocytosis; TAS.
DR InterPro: IPR002353; AntifreezeII.
DR InterPro: IPR000562; FN Type II.
DR InterPro: IPR001304; Lectin_C.
DR InterPro: IPR000772; Ricin_B_lectin.
DR Pfam: PF00040; fn2; 1.
DR Pfam: PF00059; lectin_c; 8.
DR Pfam: PF00652; Ricin_B_lectin; 2.
DR PRINTS: PR00013; PNTYPEII.
DR PRINTS: PR00356; ANTIFREEZEII.
DR ProDom: PD000995; FN_Type_II; 1.
DR SMART: SM00034; CLECT; 8.

```



```
DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; 6.
DR PROSITE; PS00041; C-TYPE LECTIN 2; 8.
DR PROSITE; PS00023; FIBRONECTIN 2; 1.
DR PROSITE; PS02031; RICIN B LECTIN; 1.
KW Receptor; Signal; Calcium-binding; Transmembrane; Repeat;
KW Glycoprotein; Endocytosis; Macrophage; Lectin; Antigen; 3D-structure.
FT SIGNAL 1 18
FT CHAIN 19 1456
FT DOMAIN 19 1456 MACROPHAGE MANNOSE RECEPTOR.
FT TRANSMEM 19 1383 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1384 1411 POTENTIAL.
FT DOMAIN 1412 1456 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 22 142 RICIN B-TYPE LECTIN.
FT DOMAIN 157 212 FIBRONECTIN TYPE-II.
FT DOMAIN 216 344 C-TYPE LECTIN 1 (LONG FORM).
FT DOMAIN 360 490 C-TYPE LECTIN 2 (LONG FORM).
FT DOMAIN 502 629 C-TYPE LECTIN 3 (LONG FORM).
FT DOMAIN 644 781 C-TYPE LECTIN 4 (LONG FORM).
FT DOMAIN 805 926 C-TYPE LECTIN 5 (LONG FORM).
FT DOMAIN 943 1083 C-TYPE LECTIN 6 (LONG FORM).
FT DOMAIN 1100 1216 C-TYPE LECTIN 7 (LONG FORM).
FT DOMAIN 1228 1359 C-TYPE LECTIN 8 (LONG FORM).
FT DISULFID 646 659
FT DISULFID 680 777
FT DISULFID 753 769
FT CARBOHYD 104 104 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 930 930 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1160 1160 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1205 1205 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1311 1311 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 648 649
FT STRAND 551 552
FT TURN 554 555
FT STRAND 558 559
FT HELIX 658 663
FT STRAND 667 669
FT STRAND 671 671
FT HELIX 673 683
FT TURN 584 584
FT STRAND 686 687
FT HELIX 693 705
FT TURN 706 707
FT TURN 709 710
FT STRAND 712 718
FT TURN 723 724
FT STRAND 727 727
FT TURN 729 730
FT STRAND 733 733
FT TURN 741 742
FT HELIX 746 748
FT STRAND 752 757
FT TURN 758 761
FT STRAND 764 768
FT TURN 769 770
FT STRAND 773 780
FT TURN 781 782
SQ SEQUENCE 1456 AA; 166011 MW; 264B5AF3C576A5E3 CRC64;

Query Match 16.3%; Score 182; DB 1; Length 1456;
Best Local Similarity 25.5%; Pred. No. 3.4e-08;
Matches 52; Conservative 37; Mismatches 61; Indels 54; Gaps 8;

QY 21 YKVIYFHTSRRLNFEAKEACRRGGQGVLSIESEDEQKLEKFIENLLPSGDFWIGLR 80
Dy 807 YKDYQYVFSKBEKTDNARAFCKRNFGLVSIQSEKKFLWKYV-NRNDQASAYFGLL 865
QY 81 RREEKQSNSTACQDLIYAWTDGSGISQFRNRYVDEPSCGS--EVCVVMYHQPSAPAGIGGPY 138
Dy 866 ISLDKK-----FAWMDGSKVDYVSWATGEFNFANEDENCVTMY-----SNSGF---- 908
```

```
139 MFOWNDRRCNNKNNFICKYSEK-----PAVPSREAGSEBETELTTPVLPEETQE----- 187
909 ---WINDINGCYPNAPICORHNSINATVMP-----TWPSVSGCKEGHNFYSN 954
188 -----EDAKYTFKESREAAAL 202
955 KCFKIFGFMEERKNWQEARAKCI 978

RESULT 4
PCGN MOUSE
ID PGCN MOUSE STANDARD; PRT; 1268 AA.
AC P55066;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
GN CSPG3 OR NCAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=96039250; PubMed=7490074;
RA Rauch U., Grimpe B., Kulbe G., Arnold-Ammer I., Beier D.,
RT Faessler R.;
RL "Structure and chromosomal localization of the mouse neurocan gene.";
RL Genomics 28:405-410(1995).
CC development by binding to neural cell adhesion molecules (NG-CAM
CC and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC acid.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X84727; CAA59216.1; -
DR PIR; S52781; S52781.
DR HSSP; P00740; LEDM.
DR MGD; MGI:104694; Cspg3.
DR InterPro; IPR002353; Antifreeze2.1.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003599; IG.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; Lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR PRINTS; PR00356; ANTIFREEZE2.1.
DR ProDom; PD000918; Link; 2.
```

DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1268
FT DOMAIN 37 157
FT DOMAIN 158 253
FT DOMAIN 259 355
FT DOMAIN 960 996
FT DOMAIN 998 1034
FT DOMAIN 1036 1165
FT DOMAIN 1166 1224
FT DISULFID 58 139
FT DISULFID 181 252
FT DISULFID 205 226
FT DISULFID 279 354
FT DISULFID 303 324
FT DISULFID 364 975
FT DISULFID 969 984
FT DISULFID 986 995
FT DISULFID 1040 1051
FT DISULFID 1068 1160
FT DISULFID 1136 1152
FT DISULFID 1167 1210
FT DISULFID 1196 1223
FT CARBOHYD 121 121
FT CARBOHYD 339 339
FT CARBOHYD 742 742
FT CARBOHYD 978 978
FT CARBOHYD 1175 1175
SQ SEQUENCE 1268 AA; 137200 MW; 30148E202A2FAC CRC64;
Query Match 15.9%; Score 177; DB 1; Length 1268;
Best Local Similarity 31.5%; Pred. No. 8.1e-08;
Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;
QY 17 ORPCYKVIYFHTSRRLNFEAEKACRRDGGQGVLSIESDEOKLIEKFTIENLLPSDGDFFW 76
Db 1048 QHCYR--YF---AHRAWEDAEKDCRRAGHLTSVHSPPEHKFINSF-----GHENSW 1096
QY 77 IGLRREERKQSNSTACQDLIYATDGSISQFRNRYVDEPS---CGSEVCVVMYHQPSAPAG 133
Db 1097 IGLNDRTVRD-----FQWTDNTGLQYENWREKQPNFFAGGEDCVVMVAHESG---- 1145
QY 134 IGGPYWFQNDRCRKNKNFICK 156
Db 1146 -----RWNDVPCVNLFPYVCK 1161
RESULT 5
PCGN_HUMAN
ID PCGN_HUMAN STANDARD; PRT; 1321 AA.
AC O14594; Q3UPK6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
GN CSFG3 OR NCAN OR NEUR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99013874; PubMed=9795216;
RA Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Lennon G.G.;
RT "Characterization of the human neurocan gene, CSPG3.";
RL Gene 221:199-205(1998).
RN [2]
RP SEQUENCE OF 1-990 AND 1007-1321 FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of an ~1 Mb region containing the MEF2B gene in
19p12.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
development by binding to neural cell adhesion molecules (NG-CAM
and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
acid.
CC -!- TISSUE SPECIFICITY: Brain.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF026547; AAC80576.1; -;
DR EMBL; AC003110; AAB86655.1; -;
DR EMBL; AC005254; AAC25581.1; -;
DR HSSP; P00740; 1EDM.
DR Genew; HGNC:2465; CSPG3.
DR MIM; 600826; -;
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_CA.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR007110; IG-Like.
DR InterPro; IPR003599; IG.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR00538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; IG; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.

```
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS0835; IG_LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lentin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1321 NEUROCAN CORE PROTEIN.
FT DOMAIN 38 153 IG-LIKE V-TYPE.
FT DOMAIN 159 254 LINK 1.
FT DOMAIN 260 356 LINK 2.
FT DOMAIN 1008 1044 EGF-LIKE 1.
FT DOMAIN 1046 1082 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1084 1213 C-TYPE LECTIN.
FT DOMAIN 1214 1272 SUSHI.
FT DISULFID 59 140 BY SIMILARITY.
FT DISULFID 182 253 BY SIMILARITY.
FT DISULFID 206 227 BY SIMILARITY.
FT DISULFID 280 355 BY SIMILARITY.
FT DISULFID 304 325 BY SIMILARITY.
FT DISULFID 1012 1023 BY SIMILARITY.
FT DISULFID 1017 1032 BY SIMILARITY.
FT DISULFID 1034 1043 BY SIMILARITY.
FT DISULFID 1088 1099 BY SIMILARITY.
FT DISULFID 1116 1208 BY SIMILARITY.
FT DISULFID 1184 1200 BY SIMILARITY.
FT DISULFID 1215 1258 BY SIMILARITY.
FT DISULFID 1244 1271 BY SIMILARITY.
FT CARBOHYD 122 122 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1223 1223 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1234 1234 N -> Y (IN REF. 2).
FT CONFLICT 1254 1254 V -> A (IN REF. 2).
FT CONFLICT 1282 1282 G -> R (IN REF. 2).
SQ SEQUENCE 1321 AA; 142972 MW; 2EF47F823DB980B8 CRC64;

Query March 15.8%; Score 176; DB 1; Length 1321;
Best Local Similarity 31.5%; Pred. NO. 1e-07;
Matches 45; Conservative 18; Mismatches 48; Indels 32; Gaps 6;

QY 17 QRCYKVIYHDTGRRLLNFEEAKACRRDGGQVLSIEDEQKLEKFIENLLPSDGDWF 76
Db 1096 QGHCYR--YF---AHRRAWEADKDCRRSHLSVHSPEHSFINSF-----GHNTW 1144

QY 77 IGLRRREKQSNACQDLYAWTQSGISQFRNYYVDEPS---CGSEVCVMYHQPAPAG 133
Db 1145 IGLNDRIVERD-----FQWTDNTGLQFNWRENQDNFFAGDCDVMVAHESG--- 1193

QY 134 IGGPYMFQWDDRCNMKNFICK 156
Db 1194 -----RWNDVPCNYPVYCK 1209

RESULT 6
PGCV_BOVIN STANDARD; PRT; 3381 AA.
ID PB1282; O77609; O77610; O77611; O77612;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP).
GN CP5G2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS VO; V1; V2 AND V3).
RC TISSUE=Forebrain;
RX MEDLINE=98288320; PubMed=9624174;
```

```
RA Schmalfeidt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
RL J. Biol. Chem. 273:15758-15764(1998).
RN [2]
RP SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331
RP AND 342-348.
RC TISSUE=Spinal cord;
RX MEDLINE=92062692; PubMed=1720020;
RA Perides G., Biviano F., Bignami A.;
RT "Interaction of a brain extracellular matrix protein with hyaluronic
RT acid.";
RL Biochim. Biophys. Acta 1075:248-258(1991).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBUNIT: Interacts with FBLN1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=P81282-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=P81282-2; Sequence=VSP_003078, VSP_003079;
CC Name=V2;
CC IsoId=P81282-3; Sequence=VSP_003080;
CC Name=V3;
CC IsoId=P81282-4; Sequence=VSP_003078, VSP_003081;
CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed
CC in the central nervous system, and in a number of mesenchymal and
CC epithelial tissues; the major isoform V2 is restricted to the
CC central nervous system.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (By similarity).
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF060456; AAC24358.1; -
CC EMBL; AF060457; AAC24359.1; -
CC EMBL; AF060458; AAC24360.1; -
CC EMBL; AF060459; AAC24361.1; -
CC PIR; T14274; T14274.
CC PIR; T42389; T42389.
CC HSP; P01132; IEPG.
CC InterPro; IPR000152; Aex_hydroxyl.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF_Like.
CC InterPro; IPR007110; Ig_Like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF000008; EGF; 2.
CC Pfam; PF00047; Ig; 1.
CC Pfam; PF00059; Lectin_c; 1.
CC Pfam; PF000084; sushi; 1.
```

DR PFam: PF00193; Xlink; 2.
 DR PRINTS; PR01265; LINKMODULE.
 DR PRINTS; PR00356; ANTIFREEZEII.
 DR PRODom; PR000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
 DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
 DR PROSITE; PS00022; EGF 1_2; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS00835; IG_LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 DR Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 20
 FT CHAIN 21 3381
 FT DOMAIN 21 147
 FT DOMAIN 168 245
 FT DOMAIN 266 347
 FT DOMAIN 349 1336
 FT DOMAIN 1337 3074
 FT DOMAIN 3074 3110
 FT DOMAIN 3112 3148
 FT DOMAIN 3161 3275
 FT DOMAIN 3280 3338
 FT DISULFID 44 131
 FT DISULFID 173 244
 FT DISULFID 197 218
 FT DISULFID 271 346
 FT DISULFID 295 316
 FT DISULFID 3078 3089
 FT DISULFID 3083 3098
 FT DISULFID 3100 3109
 FT DISULFID 3116 3127
 FT DISULFID 3121 3136
 FT DISULFID 3138 3147
 FT DISULFID 3154 3165
 FT DISULFID 3182 3274
 FT DISULFID 3250 3266
 FT DISULFID 3281 3324
 FT DISULFID 3310 3337
 FT CARBOHYD 57 57
 FT CARBOHYD 331 331
 FT CARBOHYD 352 352
 FT CARBOHYD 817 817
 FT CARBOHYD 965 965
 FT CARBOHYD 1017 1017
 FT CARBOHYD 1333 1333
 FT CARBOHYD 1393 1393
 FT CARBOHYD 1437 1437
 FT CARBOHYD 1463 1463
 FT CARBOHYD 1653 1653
 FT CARBOHYD 1974 1974
 FT CARBOHYD 2045 2045
 FT CARBOHYD 2074 2074
 FT CARBOHYD 2103 2103
 FT CARBOHYD 2263 2263
 FT CARBOHYD 2290 2290
 FT CARBOHYD 2356 2356
 FT CARBOHYD 2623 2623
 FT CARBOHYD 2641 2641
 FT CARBOHYD 2919 2919
 FT CARBOHYD 3052 3052
 FT CARBOHYD 3354 3354
 FT CARBOHYD 3364 3364
 FT VARSPLIC 349 349

FT VARSPLIC 350 1336 /FTID=VSP 003078.
 FT Missing (in isoform V1).
 FT VARSPLIC 1337 3074 /FTID=VSP 003079.
 FT Missing (in isoform V2).
 FT VARSPLIC 350 3074 /FTID=VSP 003080.
 FT Missing (in isoform V3).
 FT CONFLICT 25 25 /FTID=VSP 003081.
 FT CONFLICT 51 51 MISSING (IN REF. 2).
 FT CONFLICT 89 89 N -> D (IN REF. 2).
 FT CONFLICT 96 96 Q -> D (IN REF. 2).
 FT CONFLICT 346 346 C -> R (IN REF. 2).
 SQ SEQUENCE 3381 AA; 369984 MW; F09716FA7778D459 CRC64;
 Query Match 15.7%; Score 174.5; DB 1; Length 3381;
 1st Local Similarity 25.5%; Pred No. 4.5e-07;
 Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;
 QY 17 QRCYKVIYFHDTRRLNFEAKACRRDGGQLVSISEDEQKLEKFIENLLPSDGF- 75
 Db 3162 QGQCVK--YF--AHRRTWDAARECRLQGAHLTSLSHEEQMFVNRV-----GHDVQ 3209
 QY 76 WGLRRREEKOSNSTACODLYAWTDGSIQSRNHWYVDEP-----SCGSEVCVMYHQSAP 131
 Db 3210 WIGL-----NDKMFEDHFRWTDGSLQYENRPNQPSFFSTGDCVLIWHENG-- 3259
 QY 132 AGIGGYPWFQNDRCNNMNFICKYS---DEKPAVPSREAEGE----- 172
 Db 3260 -----QWMDVPCNVHLTYTCKGTACGQPPVVENAKTFGKMKPRYEINSLIRYHC 3310
 QY 173 -----ETELTT-----PVL-----PEETOEDAKTKFKESREAAALN 203
 Db 3311 KDGFIQRHLPTIRCLNGRWAMPKITCLNPSAYQRTYSKYPKSSSAKDN 3361
 RESULT 7
 ID PGCN RAT STANDARD; PRT; 1257 AA.
 AC P55067;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3)
 DE (245 kDa early postnatal core glycoprotein) [Contains: 150 kDa adult
 core glycoprotein].
 GN CSPG3 OR NCAN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=92406907; PubMed=1326557;
 RA Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.;
 RT "Cloning and primary structure of neurocan, a developmentally
 regulated, aggregating chondroitin sulfate proteoglycan of brain";
 RL J. Biol. Chem. 267:19536-19547(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=94230574; PubMed=7513709;
 RA Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.,
 RA Margolis R.U., Grumet M.;
 RT "The neuronal chondroitin sulfate proteoglycan neurocan binds to the
 neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits
 neuronal adhesion and neurite outgrowth.";
 RL J. Cell Biol. 125:669-680(1994).
 CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
 development by binding to neural cell adhesion molecules (NG-CAM
 and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
 acid.
 CC -!- TISSUE SPECIFICITY: EARLY POSTNATAL AND ADULT BRAIN; NOT EXPRESSED
 IN KIDNEY, LUNG, LIVER AND MUSCLE.

CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -!- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
CC DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL
CC BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M97161; AAC37679.1; -;
CC PIR; S28764; S28764.
CC HSP; P00740; IEDM.
CC InterPro: IPR002353; AntifreezeII.
CC InterPro: IPR000152; Asx hydroxyl.
CC InterPro: IPR000742; EGF 2.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR006209; EGF_Like.
CC InterPro: IPR007110; Ig_Like.
CC InterPro: IPR003599; Ig.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR001304; Lectin_C.
CC InterPro: IPR000538; Link.
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00847; Ig; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 2.
CC PRINTS; PR01265; LINKMODULE.
CC PRINTS; PR00356; ANTIFREEZEII.
CC ProDom; PD000918; Link; 2.
CC SMART; SM00032; CCP; 1.
CC SMART; SM00034; CLECT; 1.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00409; IG; 1.
CC SMART; SM00445; LINK; 2.
CC PROSITE; PS00010; ASX_HYDROXYL; 1.
CC PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
CC PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
CC PROSITE; PS00022; EGF_1; 3.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS01187; EGF_CA; 1.
CC PROSITE; PS00835; IG_LIKE; 1.
CC PROSITE; PS01241; LINK; 2.
CC KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
CC EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
CC SIGNAL 1 22
FT CHAIN 23 1257 NEUROCAN CORE PROTEIN.
FT CHAIN 639 1257 150 KDa ADULT CORE GLYCOPROTEIN.
FT DOMAIN 37 157 IG-LIKE V-TYPE.
FT DOMAIN 158 253 LINK 1.
FT DOMAIN 259 355 LINK 2.
FT DOMAIN 949 985 EGF-LIKE 1.
FT DOMAIN 987 1023 EGF-LIKE 2.
FT DOMAIN 1025 1154 C-TYPE LECTIN.
FT DOMAIN 1155 1213 SUSHI.
FT DISULFID 58 139 BY SIMILARITY.
FT DISULFID 181 252 BY SIMILARITY.
FT DISULFID 205 226 BY SIMILARITY.
FT DISULFID 279 354 BY SIMILARITY.
FT DISULFID 303 324 BY SIMILARITY.
FT DISULFID 933 964 BY SIMILARITY.

FT DISULFID 958 973 BY SIMILARITY.
FT DISULFID 975 984 BY SIMILARITY.
FT DISULFID 1029 1040 BY SIMILARITY.
FT DISULFID 1057 1149 BY SIMILARITY.
FT DISULFID 1125 1141 BY SIMILARITY.
FT DISULFID 1156 1199 BY SIMILARITY.
FT DISULFID 1185 1212 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 944 944 O-LINKED (XYL. . .) (CHONDROITIN SULFATE).
FT CARBOHYD 967 967 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1164 1164 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1257 AA; 135544 MW; 992B33DCFA19EE1B CRC64;
Query Match 15.6%; Score 174; DB 1; Length 1257;
Best Local Similarity 30.8%; Pred. No. 1.5e-07;
Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;
QY 17 QRPCKVYIFHDTSRRLNPEEAKACRRDGGQGVSTESDEQKLEKFIENLLPSDGF 76
DB 1037 QGHCYR--YF---AHRAWEDAEACRRRAGHLTSVHSPHEHKFINSF-----CHENSW 1085
QY 77 IGLRRREKQSNSTACQDIYAWTDGSIQFRNYYVDEPS---CGSEVVCVMYHQPSAPAG 133
DB 1086 IGLNDRITVERD-----FQWTDNTGLQYENREKQPDNFFAGGEDCVVMVAHENG--- 1134
QY 134 IGGPYMFQWDDRCNMKNFICK 156
DB 1135 -----RWNDVPCNYNLPYVCK 1150
RESULT 8
PGCV RAT STANDARD; PRT; 2738 AA.
ID AC Q9REB4; O08592; O08564; Q9RIK4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP) (Fragments).
GN CP5G2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.
RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wight T.N.;
RT "versican/PG-M isoforms in vascular smooth muscle cells";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [2]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=98308094; PubMed=9642104;
RA Milev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,
RA Margolis R.K., Margolis R.U.;
RT "Differential regulation of expression of hyaluronan-binding
RT proteoglycans in developing brain: aggrecan, versican, neurocan, and
RT brevican";
RL Biochem. Biophys. Res. Commun. 247:207-212(1998).
RN [3]
RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).
RC TISSUE=Kidney;
RX MEDLINE=98094159; PubMed=9434070;
RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RT "Proteoglycan expression in the normal rat kidney";
RL Nephron 77:461-470(1997).


```

RESULT 9
PCV MOUSE
ID PCV MOUSE STANDARD; PRT: 3358 AA.
AC Q62059; Q62058; Q9CU00;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
GN CSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A. (ISOFORM V0; V1 AND V2).
RC STRAIN=C57BL/6; and Swiss Webster; Tissue=Brain;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
RL J. Biol. Chem. 270:3914-3918(1995).
[3]
RN SEQUENCE OF 1-1692 FROM N.A. (ISOFORM V1).
RC STRAIN=C57BL/6J; Tissue=Skin;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guscinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[4]
RN INTERACTION WITH FBLN1.
RX PubMed=10400671;
RA Asberg A., Adam S., Kostka G., Timpl R., Heinegaard D.;
RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and
RT versican.";
RL J. Biol. Chem. 274:20444-20449(1999).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBUNIT: Interacts with FBLN1.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC isoId=Q62059-1; Sequence=Displayed;

```

```

CC Name=V1;
CC isoId=Q62059-2; Sequence=VSP_003087, VSP_003088;
CC Name=V2;
CC isoId=Q62059-3; Sequence=VSP_003089;
CC Name=V3;
CC isoId=Q62059-4; Sequence=VSP_003087, VSP_003090;
CC -!- TISSUE SPECIFICITY: V2 is found only in brain.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D16263; BAA03796.1; -
CC EMBL; D28599; -; NOT ANNOTATED_CDS.
CC EMBL; D32040; BAA06802.1; -
CC EMBL; AK014525; BAB29411.1; -
CC HSSP; P01132; IEPG.
CC MGD; MGI:102889; Cspg2.
CC InterPro; IPR000152; Asx hydroxyl.
CC InterPro; IPR000742; EGF 2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006309; EGF_Like.
CC InterPro; IPR007110; Ig_Like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00047; Ig; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 2.
CC PRINTS; PR01265; LINKMODULE.
CC ProDom; PD000918; Link; 2.
CC SMART; SM00032; CCP; 1.
CC SMART; SM00034; CLECT; 1.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00409; IG; 1.
CC SMART; SM00445; LINK; 2.
CC PROSITE; PS00010; ASX_HYDROXYL; 1.
CC PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
CC PROSITE; PS00411; C_TYPE_LECTIN_2; 1.
CC PROSITE; PS00022; EGF 1; 2.
CC PROSITE; PS01186; EGF 2; 1.
CC PROSITE; PS01187; EGF_CA; 1.
CC PROSITE; PS00835; IG_Like; 1.
CC PROSITE; PS01241; LINK; 2.
CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
CC Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
CC Hyaluronic acid; Alternative splicing.
CC SIGNAL 1 20 POTENTIAL,
CC CHAIN 21 3358 VERSICAN CORE PROTEIN.
CC DOMAIN 21 146 IG-LIKE V-TYPE.
CC DOMAIN 167 244 LINK 1.
CC DOMAIN 265 346 LINK 2.
CC DOMAIN 348 1308 GAG-ALPHA
CC (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
CC DOMAIN 1309 3052 GAG-BETA.
CC DOMAIN 3052 3088 EGF-LIKE 1.
CC DOMAIN 3090 3126 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
CC DOMAIN 3139 3253 C-TYPE LECTIN.

```

```

FT DOMAIN 3258 3316 SUSHI.
FT DISULFID 44 130 BY SIMILARITY.
FT DISULFID 172 243 BY SIMILARITY.
FT DISULFID 196 217 BY SIMILARITY.
FT DISULFID 270 333 BY SIMILARITY.
FT DISULFID 294 315 BY SIMILARITY.
FT DISULFID 306 367 BY SIMILARITY.
FT DISULFID 3061 3076 BY SIMILARITY.
FT DISULFID 3078 3087 BY SIMILARITY.
FT DISULFID 3094 3105 BY SIMILARITY.
FT DISULFID 3099 3114 BY SIMILARITY.
FT DISULFID 3116 3125 BY SIMILARITY.
FT DISULFID 3112 3143 BY SIMILARITY.
FT DISULFID 3160 3252 BY SIMILARITY.
FT DISULFID 3228 3244 BY SIMILARITY.
FT DISULFID 3259 3302 BY SIMILARITY.
FT DISULFID 3288 3315 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 914 914 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1305 1305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1372 1372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1679 1679 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2054 2054 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2244 2244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2362 2362 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2627 2627 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3030 3030 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3332 3332 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3342 3342 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 348 348 P -> R (in isoform V1 and isoform V3).
FT VARSPLIC 349 1308 Missing (in isoform V1).
FT VARSPLIC 1309 3052 Missing (in isoform V2).
FT VARSPLIC 349 3052 Missing (in isoform V3).
FT CONFLICT 126 126 A -> G (IN REF. 3).
FT CONFLICT 348 348 MISSING (IN REF. 3).
FT CONFLICT 1658 1658 I -> T (IN REF. 3).
FT CONFLICT 1674 1680 TVMNSNS -> QFGIQTA (IN REF. 3).
SQ SEQUENCE 3358 AA; 366938 MW; 071B80026BC0762D CRC64;

Query Match 15.68; Score 174; DB 1; Length 3358;
Best Local Similarity 28.58; Pred. No. 5e-07;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

Qy 17 QRPCYKVIYFHTSRRLLNFEAEKAEARRDGGQLVSIIESEDEKLIEXFIENLLPSDGDGF- 75
Db 3140 QGCYK--YF--AHRRTWDAARECLQNAHLTSLSHQEMFVNRV-----GHDIQ 3187

Qy 76 WIGLRREEKQSNSTACQDIYATDGSISQFRNMYVDEP----SCGSEVCVMYHQPSAP 131
Db 3188 WIGL-----NDKMEPHDFRWTGSLQVYENRPNQPSFFTSAGEDCVVIWHENG-- 3237

Qy 132 AGIGPPYMFQNDRCNKNKFNFTCKYS----DEKPAVPREASGE 172
Db 3238 -----QWNDVPCNHYLTCTCKGTACGPPVVENAKTFGK 3273

RESULT 10
PGCV_HUMAN
ID_PGCV_HUMAN STANDARD; PRT; 3396 AA.
AC P1361; P20754; Q13010; Q13189; Q15123; Q9UNW5;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)

```

```

DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP).
GN CSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORM V0).
RX MEDLINE=95105186; PubMed=7528742;
RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
RT "Characterization of the complete genomic structure of the human
RT versican gene and functional analysis of its promoter.";
RL J. Biol. Chem. 269:32999-33008(1994).
RN (2)
RP SEQUENCE FROM N.A. (ISOFORM V1).
RX TISSUE=Placenta;
RA MEDLINE=90059882; PubMed=2581089;
RA Zimmermann D.R., Ruoslahti E.;
RT "Multiple domains of the large fibroblast proteoglycan, versican.";
RL EMBO J. 8:2975-2981(1989).
RN (3)
RP SEQUENCE FROM N.A. (ISOFORM V2).
RX TISSUE=Glial tumor;
RA MEDLINE=95105187; PubMed=7806529;
RA Dours-Zimmermann M.T., Zimmermann D.R.;
RT "A novel glycosaminoglycan attachment domain identified in two
RT alternative splice variants of human versican.";
RL J. Biol. Chem. 269:32992-32998(1994).
RN (4)
RP SEQUENCE OF 2711-3396 FROM N.A.
RX TISSUE=Lung fibroblast;
RA MEDLINE=88007514; PubMed=2820964;
RA Krusius T., Gehlsen K.R., Ruoslahti E.;
RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
RT lectin-like and growth factor-like sequences.";
RL J. Biol. Chem. 262:13120-13125(1987).
RN (5)
RP SEQUENCE OF 251-347 FROM N.A.
RX MEDLINE=93122792; PubMed=1478664;
RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
RA Meperson J.D.;
RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
RT human chromosome 5 (5q12-5q14).";
RL Genomics 14:845-851(1992).
RN (6)
RP SEQUENCE FROM N.A. (ISOFORM V3).
RX TISSUE=Brain;
RA MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
RL J. Biol. Chem. 270:3914-3918(1995).
RN (7)
RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINTI).
RX TISSUE=Aortic smooth muscle;
RA MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wight T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN (8)
RP PARTIAL SEQUENCE.
RX TISSUE=Brain;
RA MEDLINE=89174663; PubMed=2466833;
RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
RT "Isolation and partial characterization of a glial
RT hyaluronate-binding protein.";
RL J. Biol. Chem. 264:5981-5987(1989).
RN (9)
RP TISSUE SPECIFICITY OF ISOFORMS.
RX MEDLINE=96213482; PubMed=8627343;

```


RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
RT "Differential expression of versican isoforms in brain tumors.";
RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBUNIT: Interacts with FBLN1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=PI3611-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=PI3611-2; Sequence=VSP_003082, VSP_003083;
CC Name=V2;
CC IsoId=PI3611-3; Sequence=VSP_003084;
CC Name=V3;
CC IsoId=PI3611-4; Sequence=VSP_003082, VSP_003085;
CC Name=Vint;
CC IsoId=PI3611-5; Sequence=VSP_003086;
CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
CC in normal brain, gliomas, medulloblastomas, schwannomas,
CC neurofibromas, and meningiomas; V2 is restricted to normal brain
CC and gliomas; V3 is found in all these tissues except
CC medulloblastomas.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U16306; AA65018.1; -;
CC EMBL; X15998; CA34128.1; -;
CC EMBL; S52488; AA824878.1; -;
CC EMBL; U26555; AA67565.1; -;
CC EMBL; D32039; BAA06801.1; -;
CC EMBL; J02814; AAA36437.1; -;
CC EMBL; AF084545; AAD48545.1; -;
CC PIR; S06014; A60979.
CC HSSP; P01132; IEGF.
CC Genew; HGNC:2464; CSPG2.
CC MIM; 118661; -;
CC GO; GO:0005578; C:extracellular matrix; TAS.
CC GO; GO:0005204; F:chondroitin sulfate proteoglycan; TAS.
CC GO; GO:0005540; F:hyaluronic acid binding activity; TAS.
CC GO; GO:0008037; P:cell recognition; TAS.
CC GO; GO:0007275; P:development; TAS.
CC InterPro; IPR000152; Asx_hydroxyl.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF_Like.
CC InterPro; IPR007110; IG_Like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 2.

DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS50835; IG_Like; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 3396
FT DOMAIN 21 146
FT DOMAIN 167 244
FT DOMAIN 265 346
FT DOMAIN 348 1335
FT DOMAIN 1336 3089
FT DOMAIN 3089 3125
FT DOMAIN 3127 3163
FT DOMAIN 3176 3290
FT DOMAIN 3295 3353
FT DISULFID 44 130
FT DISULFID 172 243
FT DISULFID 196 217
FT DISULFID 270 345
FT DISULFID 294 315
FT DISULFID 3093 3104
FT DISULFID 3098 3113
FT DISULFID 3115 3124
FT DISULFID 3131 3142
FT DISULFID 3136 3151
FT DISULFID 3153 3162
FT DISULFID 3169 3180
FT DISULFID 3197 3289
FT DISULFID 3265 3281
FT DISULFID 3296 3339
FT DISULFID 3325 3352
FT CARBOHYD 57 57
FT CARBOHYD 330 330
FT CARBOHYD 615 615
Query Match 15.6%; Score 174; DB 1; Length 3396;
Best Local Similarity 28.5%; Pred. No. 5.1e-07;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
QY 17 QPCYKVIYHDTSRRLNFEEAKEARRDGGQGLVSISEDEQKLIETIENLLPSDGDGF- 75
Db 3177 QGQCYK-YF---AHRRTWDAARECRLOGAHLTSLSHSEQMFVNRV-----GHDVQ 3274
QY 76 WIGLRREEKQSNSTACQDLYAWTDGSIQSRNRYVDEP-----SCGSEVCVVMYHQPAP 131
Db 3225 WIGL-----NDKMFEDHDFRMTDGTLOYNRPNQPDSPFFSAGEDCVVIWHENG-- 3274
QY 132 AGIGPYMFWQNDDRCKMKNFKICKYS-----DEKPAVPSRAEGE 172
Db 3275 -----QNDVPCNHLTYTCKGTVACGQPPVVENAKTFGK 3310
RESULT 11
PGCV CHICK
ID -PGCV CHICK STANDARD; PRT; 3562 AA.
AC Q90953; Q90945;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Versican core protein precursor (Large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
 CN CP52.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1].
 RP SEQUENCE FROM N.A. (ISOFORMS V0 AND V1).
 RC STRAIN=White leghorn; TISSUE=Limb bud;
 RA MEDLINE=93300846; PubMed=8314802;
 RX Shinomura T., Nishida Y., Ito K., Kimata K.;
 RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
 RT expressed during chondrogenesis in chick limb buds. Alternative
 RT spliced multiforms of PG-M and their relationships to versican.";
 RL J. Biol. Chem. 268:14461-14469(1993).
 CC -!- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=V0;
 CC IsoId=Q90953-1; Sequence=Displayed;
 CC Name=V1;
 CC IsoId=Q90953-2; Sequence=VSP_003093;
 CC -!- TISSUE SPECIFICITY: Prechondrogenic condensation area of
 CC developing limb buds.
 CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
 CC (By similarity).
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X60226; CAA42787.1; -;
 DR EMBL; D13542; BAA02742.1; -;
 DR PIR; A47171; A47171.
 DR HSSP; P00740; 1EDM.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR007110; Ig_Like.
 DR InterPro; IPR003599; Ig_MHC.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00047; ig; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.
 DR PRINTS; PR01265; LINKMODULE.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
 DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS00835; IG_Like; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 26
 FT CHAIN 27 3562
 FT DOMAIN 27 143
 FT DOMAIN 166 243
 FT DOMAIN 264 345
 FT DOMAIN 325 329
 FT DOMAIN 329 332
 FT DOMAIN 332 336
 FT DOMAIN 336 341
 FT DOMAIN 341 345
 FT DOMAIN 345 3518
 FT DISULFID 44 129
 FT DISULFID 171 242
 FT DISULFID 195 216
 FT DISULFID 269 344
 FT DISULFID 293 314
 FT DISULFID 325 329
 FT DISULFID 329 332
 FT DISULFID 332 336
 FT DISULFID 336 341
 FT DISULFID 341 345
 FT DISULFID 345 3518
 FT CARBOHYD 163 163
 FT CARBOHYD 235 235
 FT CARBOHYD 329 329
 FT CARBOHYD 529 529
 FT CARBOHYD 709 709
 FT CARBOHYD 948 948
 FT CARBOHYD 1409 1409
 FT CARBOHYD 1479 1479
 FT CARBOHYD 1523 1523
 FT CARBOHYD 1530 1530
 FT CARBOHYD 1625 1625
 FT CARBOHYD 1751 1751
 FT CARBOHYD 1988 1988
 FT CARBOHYD 2088 2088
 FT CARBOHYD 2089 2089
 FT CARBOHYD 2507 2507
 FT CARBOHYD 2642 2642
 FT CARBOHYD 2679 2679
 FT CARBOHYD 2748 2748
 FT CARBOHYD 2762 2762
 FT CARBOHYD 3069 3069
 FT CARBOHYD 3194 3194
 FT CARBOHYD 3232 3232
 FT CARBOHYD 3545 3545
 FT VARSPLIC 485 1411
 FT SQ SEQUENCE 3562 AA; 388078 MW; 9BC565E88C1602D2 CRC64;
 Query Match 15.3%; Score 171; DB 1; Length 3562;
 Best Local Similarity 28.5%; Pred. No. 9.9e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
 QY 17 QRCYKVIYFHTDTSRRLLNFEAKEACRDCGQVLSIESEDEOKLIEFIENLLPSDGF- 75
 Db 3342 QGOCYK--YF---AHRRTWTAERECKLQGAHLTSLUSHEEQVFNRI-----GHDYQ 3389
 QY 76 WIGLRREKQSNSTACQDLVYAWTDGSIQFRNMYVDEP-----SCGSEVCVMYHQPSP 131

```

Db 3390 WIGL-----NDKMFDRFWTDGSPLOYENRPNQDPSFSGEDCVIIWHENG-- 3439
QY 132 AGIGPYMFWNDNCKNNFKICKYS-----DEKPAVPSREAGE 172
Db 3440 -----QMDVPCNVHLTYTCKKGTVACQGPVVENAKTFGK 3475

RESULT 12
CD93_RAT
ID CD93_RAT STANDARD; PRT; 643 AA.
AC Q9E161; Q9J126;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Complement component C1q receptor precursor (Complement component 1, q
DE subcomponent, receptor 1) (ClqR) (ClqR(p)) (Clq/MBL/SPA receptor)
DE (CD93 antigen) (Cell surface antigen AA4).
DE ClqR1 OR CD93 OR ClqRP.
GN Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=PVG; TISSUE=Natural killer cells;
RX MEDLINE=20545218; PubMed=11093152;
RA Lovik G., Vaage J.T., Dissen E., Szpirer C., Ryan J.C., Rolstad B.;
RT "Characterization and molecular cloning of rat ClqR, a receptor on NK
RT cells.";
RL Eur. J. Immunol. 30:3355-3362(2000).
RN [2]
SEQUENCE FROM N.A.
RX STRAIN=Wistar; TISSUE=Lung;
RX MEDLINE=20507883; PubMed=10934210;
RA Dean Y.D., McGreal E.P., Akatsu H., Gasque P.;
RT "Molecular and cellular properties of the rat AA4 antigen, a C-type
RT lectin-like receptor with structural homology to thrombomodulin.";
RL J. Biol. Chem. 275:34382-34392(2000).
CC -!- FUNCTION: Receptor for element of a larger receptor complex) for
CC Clq, mannose-binding lectin (MBL2) and pulmonary surfactant
CC protein A (SPA). May mediate the enhancement of phagocytosis in
CC monocytes and macrophages upon interaction with soluble defense
CC collagens. May play a role in intercellular adhesion.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Widely expressed. Highly expressed in lung and
CC heart. Expressed at lower level in brain, thymus, liver, spleen,
CC intestine, kidney, adrenal gland, muscle and testis. Expressed on
CC endothelial cells, platelets, undifferentiated monocytes and
CC circulating natural killer cells.
CC -!- PTM: N- and O-glycosylated (By similarity).
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 5 EGF-like domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AF136537; RAG01572.1; -
DR EMBL; AF160978; RAG0402.1; -
DR HSP; P35555; 1EMN.
DR GO; GO:0016021; C:integral to membrane; ISS.
DR GO; GO:0004872; F:receptor activity; ISS.
DR GO; GO:0016337; P:cell-cell adhesion; ISS.
DR GO; GO:0042116; P:macrophage activation; ISS.
DR GO; GO:0006909; P:phagocytosis; ISS.
DR InterPro; IPR00152; Asx hydroxyl.
DR InterPro; IPR001881; EGF_Ca
DR InterPro; IPR006209; EGF_Like.

```

```

DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00008; EGF; 5.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 3.
KW Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
EGF-like domain; Lectin; Glycoprotein.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 643 COMPLEMENT COMPONENT C1Q RECEPTOR.
FT DOMAIN 24 571 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 572 592 POTENTIAL.
FT DOMAIN 593 643 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 31 173 C-TYPE LECTIN.
FT DOMAIN 257 298 EGF-LIKE 1.
FT DOMAIN 299 341 EGF-LIKE 2.
FT DOMAIN 342 381 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 382 423 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 424 462 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 268 282 BY SIMILARITY.
FT DISULFID 284 297 BY SIMILARITY.
FT DISULFID 303 314 BY SIMILARITY.
FT DISULFID 308 325 BY SIMILARITY.
FT DISULFID 327 340 BY SIMILARITY.
FT DISULFID 346 355 BY SIMILARITY.
FT DISULFID 351 364 BY SIMILARITY.
FT DISULFID 366 380 BY SIMILARITY.
FT DISULFID 386 397 BY SIMILARITY.
FT DISULFID 393 406 BY SIMILARITY.
FT DISULFID 408 422 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 443 446 BY SIMILARITY.
FT DISULFID 448 461 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 498 498 E -> K (IN REF. 2).
FT CONFLICT 417 417
SQ SEQUENCE 643 AA; 68781 MW; 9AE4C933AD943DB6 CRC64;
Query Match 14.8%; Score 165; DB 1; Length 643;
Best Local Similarity 25.2%; Pred. No. 4.2e-07;
Matches 52; Conservative 35; Mismatches 73; Indels 46; Gaps 10;
QY 3 GRLLSGQP-----VCRGRTORPCYKVIYFHDTSRLNPFEEAKEACRRDGGQLVSI ESD 56
Db 15 GOLWAGAAADSEAVVCEG---TACYTAHW-----GKLSAAEAQHRCNENGSLATVKSEE 66
QY 57 EOKLIEKFIENLL-----PSD---GDFWICLRREEKQSNSTACQDLYAWT-DGSISQFRN 108
Db 67 EARHVOEALAOQLTKKAPSETKIGKFWIGLOREKGTVHDLPMKGFWSVGGEEDTYSN 126
QY 109 WY-VDEPSCGSVCVVMV-----HQSAPAGIGPYMFWNDDRD-----NMKNPF 153
Db 127 WYKASKSCISKRCVSLILDLSLKHPSHLP-----KWHESPCTGTPDAGNSIEGF 177
QY 154 ICKYSDKPAVPSREAGEETLTPP 179
Db 178 LCKFNFKMGKCSPLALGPGQLTYTTP 203
RESULT 13
PGCA_BOVIN
ID PGCA_BOVIN STANDARD; PRT; 2364 AA.
AC P13608; P79117; Q28159;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSPCP).

```



```
FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 509 580 BY SIMILARITY.
FT DISULFID 533 554 BY SIMILARITY.
FT DISULFID 607 682 BY SIMILARITY.
FT DISULFID 631 652 BY SIMILARITY.
FT DISULFID 2117 2128 BY SIMILARITY.
FT DISULFID 2250 2266 BY SIMILARITY.
FT DISULFID 2281 2324 BY SIMILARITY.
FT DISULFID 2310 2337 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 2114 2150 Missing (in isoform 2).
FT SEQUENCE 2364 AA; 246359 MW; 6FF83763420C3D4C CRC64;

Query Match 14.2%; Score 158.5; DB 1; Length 2364;
Best Local Similarity 26.8%; Pred. No. 7.8e-06;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;

QY 9 QVCRGG---TORPCYKVIYFHTSRRLNFEBAKEACRRDGGQLVSISEDEQKLIEXPI 65
Db 2151 QKLCBEGWTKFGHCYR--HFPD---RATWVDAESQCRKQKQSHLSSIVTPERQ----EFV 2201

QY 66 ENLLPSDGF-WGLRRREEKOSNSTACQDLYAWTDGSIQFRNMYVDEP-----SCGSEV 120
Db 2202 NN---NAQYQWIGL-----NKTTEGDFRWDGHSLOPENRPNQPDNFFATGEDC 2250

QY 121 CVMYHQPASAPAGIGPYFMQNDRCNMKNFICKYS----DEKPAVPSREAGEETE 175
Db 2251 VVMWHEKG-----EWNDVPCNVQLPFTCKGTGTVACGEPVVEHARIFQOKD 2298

RESULT 14
PCGA HUMAN STANDARD; PRT; 2415 AA.
AC P16112; O13650;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSPCP) (Chondroitin sulfate proteoglycan core protein 1).
GN AGC1 OR CSPG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Chondrocytes;
RX MEDLINE=91093289; PubMed=1985970;
RA Doegi K.J., Sasaki M., Kimura T., Yamada Y.;
RT "Complete coding sequence and deduced primary structure of the human
RT cartilage large aggregating proteoglycan, aggrecan. Human-specific
RT repeats, and additional alternatively spliced forms.";
RL J. Biol. Chem. 266:894-902(1991).
RN [2]
RP SEQUENCE OF 1778-2415 FROM N.A. (ISOFORM 2).
RC TISSUE=Chondrocytes;
RA Dudhia J., Hardingham T.E.;
RL Submitted (JAN-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1936-2415 FROM N.A. (ISOFORM 1).
RX MEDLINE=89380154; PubMed=2789216;
RA Baldwin C.T., Reginato A.M., Prockop D.J.;
RT "A new epidermal growth factor-like domain in the human core protein
RT for the large cartilage-specific proteoglycan. Evidence for
RT alternative splicing of the domain.";
RL J. Biol. Chem. 264:15747-15750(1989).
```

```
[4]
RN RP SEQUENCE OF 764-864 FROM N.A.
RX TISSUE=Blood;
RC MEDLINE=95128522; PubMed=7827755;
RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
RT "Length variation in the keratan sulfate domain of mammalian
RT aggrecan.";
RL Matrix Biol. 14:323-328(1994).
CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION.
CC -!- SUBUNIT: Interacts with FBLN1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=PI61112-1; Sequence=VSP_003074;
CC Name=2;
CC IsoId=PI61112-2; Sequence=VSP_003074;
CC Name=3;
CC IsoId=PI61112-3; Sequence=VSP_003074, VSP_003075;
CC -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 4 link domains.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M55172; AAA62824.1; -
CC EMBL; J05062; AAA35726.1; -
CC EMBL; X17406; CAA35463.1; -
CC EMBL; S74659; AAC60643.2; -
CC PIR; A39086; A39086.
CC HSP; P98066; I1SG.
CC Genew; HGNC:319; AGC1.
CC MIM; 155760; -
CC GO; GO:0005204; F:chondroitin sulfate proteoglycan; TAS.
CC InterPro; IPR002353; AntifreezeII.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR003324; SGXSG.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00047; Ig; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF02339; SGXSG; 71.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 4.
```


| Result No. | Score | Query % | | DB | ID | Description |
|------------|-------|---------|--------|----|--------|---------------------|
| | | Match | Length | | | |
| 1 | 182 | 16.3 | 1456 | 1 | A3563 | mannose receptor p |
| 2 | 178.5 | 16.0 | 1455 | 1 | A4925 | mannose receptor p |
| 3 | 177 | 15.9 | 1368 | 2 | S52781 | neurocan - mouse |
| 4 | 174.5 | 15.7 | 1643 | 2 | T14274 | versican precursor |
| 5 | 174.5 | 15.7 | 3381 | 2 | T42389 | versican precursor |
| 6 | 174 | 15.6 | 1257 | 2 | S28764 | neurocan precursor |
| 7 | 174 | 15.6 | 2397 | 1 | A55535 | versican precursor |
| 8 | 174 | 15.6 | 2409 | 1 | A60979 | versican precursor |
| 9 | 171 | 15.3 | 3562 | 2 | A47171 | chondroitin sulfat |
| 10 | 170.5 | 15.3 | 1479 | 2 | T42710 | mannose receptor, |
| 11 | 158.5 | 14.2 | 1340 | 2 | A39808 | proteoglycan core |
| 12 | 158.5 | 14.2 | 2327 | 2 | T42630 | aggreccan - bovine |
| 13 | 158.5 | 14.2 | 2115 | 1 | A39086 | aggreccan - bovine |
| 14 | 154.5 | 13.9 | 612 | 2 | B42755 | aggreccan precursor |
| 15 | 153.5 | 13.8 | 2124 | 2 | A28452 | E-selectin precurs |
| 16 | 152 | 13.6 | 912 | 2 | A54423 | proteoglycan core |
| 17 | 149.5 | 13.4 | 459 | 2 | T24425 | brevican precursor |
| 18 | 149 | 13.4 | 321 | 1 | LNHUER | hypothetical prote |
| 19 | 148.5 | 13.3 | 330 | 2 | T46256 | IGF FC receptor II |
| 20 | 148.5 | 13.3 | 2132 | 1 | A55182 | brevican - human |
| 21 | 148 | 13.3 | 253 | 2 | E89130 | aggreccan precursor |
| 22 | 147 | 13.2 | 883 | 2 | S97653 | protein F52E1.2 (i |
| 23 | 146.5 | 13.1 | 162 | 1 | LNRC1 | brevican precursor |
| 24 | 146 | 13.1 | 2109 | 1 | I50421 | lectin BRA3-1 prec |
| 25 | 145.5 | 13.0 | 742 | 2 | JC7595 | aggreccan precursor |
| 26 | 145 | 13.0 | 883 | 2 | S49126 | scavenger receptor |
| 27 | 144.5 | 13.0 | 173 | 2 | S10548 | brevican precursor |
| 28 | 144.5 | 13.0 | 372 | 2 | S23936 | lectin - barnacle |
| 29 | 144.5 | 13.0 | 404 | 2 | A46274 | L-selectin precurs |
| 30 | 144.5 | 13.0 | 404 | 2 | A46274 | HIV gp120-binding |

```
Db 807 YKDYQYFSEKETMDNARAFCKNFGLVSIQSEKFLWKYV-NRNDQAQSYFICLL 865
Qy 81 RREEKQSNSTACQDLVATWDTGSIQFRNYYVDEPSCGS--EVCVMYHQPSAPAGIGPY 138
Db 866 ISLDKK-----FAMWDSKYVYSWATGENFANEDENCVTMY-----SNSGF----- 908
Qy 139 MFQWDDRCNMKNFICKYSEK-----PAVPSREAEGETELTTPVLPEETQE----- 187
Db 909 ---WNDINCYPNFAFCQRHNSINATTVMP-----TWPSVPSGCKEWMFYSN 954
Qy 188 -----EDAKKTFKESREAAAL 202
Db 955 KCFKIFGFMEEERKNWQEARAKCI 978

RESULT 2
A48925
mannose receptor precursor, macrophage - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A48925; S21320; PC2245
R:Harris, N.; Super, M.; Rits, M.; Chang, G.; Ezekowitz, R.A.
Blood 80, 2363-2373, 1992
A:Title: Characterization of the murine macrophage mannose receptor: demonstration that
A:Reference number: A48925; MUID:93043353; PMID:1421407
A:Accession: A48925
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1455 <HA2>
A:Experimental source: peritoneal macrophage
A>Note: sequence extracted from NCBI backbone (NCBIP:118733)
R:Harris, N.; Super, M.; Rits, M.; Chang, G.; Ezekowitz, R.B.
submitted to the EMBL Data Library, April 1992
A:Description: Characterization of the murine macrophage mannose receptor: Demonstration
on.
A:Reference number: S21320
A:Accession: S21320
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-302, 'W', 303-1117, 'E', 1119-1455 <HA2>
A:Cross-references: EMBL:Z11974; NID:952997; PIDN:CAA78028.1; PID:952998
R:Harris, N.; Peters, L.L.; Eicher, E.M.; Rits, M.; Raspberry, D.; Eichbaum, Q.G.; Super
Biochem. Biophys. Res. Commun. 198, 682-692, 1994
A:Title: The exon-intron structure and chromosomal localization of the mouse macrophage
A:Reference number: PC2245; MUID:94128116; PMID:8297379
A:Accession: PC2245
A:Molecule type: mRNA
A:Residues: 35-105 <HA3>
C:Genetics:
A:Gene: Mrcl
A:Map position: 2
C:Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II re
C:Keywords: membrane protein; receptor
F:168-209/Domain: fibronectin type II repeat homology <2F9>
F:161-485/Domain: C-type lectin homology <LCH1>
F:943-1077/Domain: C-type lectin homology <LCH2>

Query Match 16.0%; Score 178.5; DB 1; Length 1455;
Best Local Similarity 25.9%; Pred. No. 1.8e-07;
Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;

Qy 21 YKVIYFHTSRLNFEAEKACRRDGGQGVLSIASEDEQKLIKFLENLPSDGFWIGLR 80
Db 806 YKDYQYFSEKETMDNARAFCKNFGLVSIQSEKFLWKYI-NKNGQGSYPYFIMGL 864
Qy 81 RREEKQSNSTACQDLVATWDTGSIQFRNYYVDEPSCGS--EVCVMYHQPSAPAGIGPY 138
Db 865 ISWDXK-----FIWMDGSKVDFVAMATGENFANEDENCVTMY-----TNSGF----- 907
Qy 139 MFQWDDRCNMKNFICK---YSDEKPAVPSREAEGETELTTPVLPEETQE----- 187
Db 908 ---WNDINCYPNFAFCQRHNSINATAMP-----TTPTFGCKEKGWHLKYNK 953
```

```
Qy 188 -----EDAKKTFKESREAAALNL 204
Db 954 CFKIFGPFANEKSKWQDARQACKGL 978
```

RESULT 3

```
S52781
neurocan - mouse
C:Species: Mus musculus (house mouse)
C>Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 04-Feb-2000
C:Accession: S52781
R:Rauch, U.; Forberg, N.; Kuibe, G.; Arnold-Ammer, I.; Faessler, R.
submitted to the EMBL Data Library, February 1995
A:Description: Amino acid sequence of mouse neurocan and brevicin and their different
A:Reference number: S52781
A:Accession: S52781
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1268 <RAU>
A:Cross-references: EMBL:X84727; NID:9758629; PIDN:CAA59216.1; PID:9758630
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
F:176-453/Domain: link protein repeat homology <LNK1>
F:274-453/Domain: link protein repeat homology <LNK2>
F:964-995/Domain: EGF homology <EGF>
F:1040-1160/Domain: C-type lectin homology <LCH>
F:1167-1223/Domain: complement factor H repeat homology <FHD>
```

```
Query Match 15.9%; Score 177; DB 2; Length 1268;
Best Local Similarity 31.5%; Pred. No. 2.1e-07;
Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;
```

```
Qy 17 QRPCYKVIYFHTSRLNFEAEKACRRDGGQGVLSIASEDEQKLIKFLENLPSDGF 76
Db 1048 QGHCYR--YF---AHRWMDAEDRCRRRAGHTSVHSPEHKPFINSF-----GHNSW 1096
```

```
Qy 77 IGLRREEKQSNSTACQDLVATWDTGSIQFRNYYVDEPS---CGSEVVCVMYHQPSAPAG 133
Db 1097 IGLNDRTVRD-----FQWTDNTGLQYENNRKQPDNFFAGGEDCVVVAHESG--- 1145
```

```
Qy 134 IGGPYMFQWDDRCNMKNFICK 156
```

```
Db 1146 -----RWNDVPCNYPYVCK 1161
```

RESULT 4

```
T14274
versican precursor, splice form V2 - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000
C:Accession: T14274
R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
J. Biol. Chem. 273, 15758-15764, 1998
A:Title: Versican V2 is a major extracellular matrix component of the mature bovine bra
A:Reference number: Z17954; MUID:98288320; PMID:9624174
A:Accession: T14274
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1643 <SCH>
A:Cross-references: EMBL:AF060458; NID:93253303; PID:93253304; PIDN:AAC24360.1
A:Experimental source: brain
C:Keywords: glycoprotein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-1643/Product: versican, splice form V2 #status predicted <MAT>
F:57,331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (Asn) (covalent) #
```

```
Query Match 15.7%; Score 174.5; DB 2; Length 1643;
Best Local Similarity 25.5%; Pred. No. 4.8e-07;
Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;
```

```
Qy 17 QRPCYKVIYFHTSRLNFEAEKACRRDGGQGVLSIASEDEQKLIKFLENLPSDGF 75
Db 1424 QGQCYK--YF---AHRRTWDAARECRQLQCAHLTSLSHEEQMFVNRV-----GHDYQ 1471
```


RESULT 8
A60979
N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan
N:Contains: glial hyaluronate-binding protein
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: S06014; S43921; A60979; A30358; A29348; A45131; I54179
R:Zimmermann, D.R.; Ruoslahti, E.
EMBO J. 8, 2975-2981, 1989
A:Title: Multiple domains of the large fibroblast proteoglycan, versican.
A:Reference number: S06014; MUID:90059882; PMID:2583089
A:Accession: S06014
A:Molecule type: mRNA
A:Residues: 1-2409 <ZIM>
A:Cross-references: GB:X15998; NID:g37662; PIDN:CAA34128.1; PID:g37663
R:Yao, L.Y.; Moody, C.; Schoenher, E.; Wight, T.N.; Sandell, L.J.
Matrix Biol. 14, 213-225, 1994
A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by
A:Reference number: S43921; MUID:95005762; PMID:7921538
A:Accession: S43921
A:Molecule type: mRNA
A:Residues: 208-440; 1094-1385; 1910-2246 <YAO>
R:Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.
Brain Res. Bull. 22, 67-70, 1989
A:Title: Structural similarity of hyaluronate binding proteins in brain and cartilage.
A:Reference number: A60979; MUID:89229983; PMID:2469524
A:Accession: A60979
A:Molecule type: protein
A:Residues: 171-210; 289-303 <BIG>
R:Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.
J. Biol. Chem. 264, 5981-5987, 1989
A:Title: Isolation and partial characterization of a glial hyaluronate-binding protein.
A:Reference number: A30358; MUID:89174663; PMID:2466833
A:Accession: A30358
A:Molecule type: protein
A:Residues: 24-80; 87-107; 119-128; 155-167; 218-229; 259-261; 268-277-283, 'G', 285-
R:Krusius, T.; Gehlsen, K.R.; Ruoslahti, E.
J. Biol. Chem. 262, 13120-13125, 1987
A:Title: A fibroblast chondroitin sulfate proteoglycan core protein contains lectin-like
A:Reference number: A29348; MUID:88007514; PMID:2820964
A:Accession: A29348
A:Molecule type: mRNA
A:Residues: 1725, 'V', 1727-2409 <KRU>
A:Cross-references: GB:J02814
R:Perides, G.; Rahemtulla, F.; Lane, W.S.; Asher, R.A.; Bignami, A.
J. Biol. Chem. 267, 23883-23887, 1992
A:Title: Isolation of a large aggregating proteoglycan from human brain.
A:Reference number: A45131; MUID:93054750; PMID:1429726
A:Contents: brain
A:Accession: A45131
A:Molecule type: protein
A:Residues: 21-22, 'X', 24-37 <PE2>
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIP:118884)
R:Iozzo, R.V.; Naso, M.F.; Cannizzaro, L.A.; Wasmuth, J.J.; McPherson, J.D.
Genomics 14, 845-851, 1992
A:Title: Mapping of the versican proteoglycan gene (CSPG2) to the long arm of human chro
A:Reference number: I54179; MUID:93122792; PMID:1478664
A:Accession: I54179
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 251-347 <RES>
A:Cross-references: GB:S52408; NID:g263313; PIDN:AAB24878.1; PID:g263314
C:Genetics:
A:Gene: GDB:CSPG2
A:Cross-references: GDB:127873; OMIM:118661
A:Map position: Sq12-Sq14
C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EG
F:1-20/Domain: signal sequence #status predicted <SIG>
F:121-2409/Product: proteoglycan 24K core protein #status predicted <MAT>
F:167-244/Domain: link protein repeat homology <LNK1>

F:265-346/Domain: link protein repeat homology <LNK2>
F:539-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>
F:2106-2137/Domain: EGF homology <EG1>
F:2144-2175/Domain: EGF homology <EG2>
F:2182-2302/Domain: C-type lectin homology <LCH>
F:2309-2365/Domain: complement factor H repeat homology <FHD>
Query Match 15.6%; Score 174; DB 1; Length 2409;
Best Local Similarity 28.5%; Pred. No. 8.3e-07;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
QY 17 QRPCKVIYFHDTSRLNFEFEAKEACRRDGGQVLSIESEDEQKLEKFIENLLPSDGF- 75
DB 2190 QGQCYK--YF---AHRRTWDAARECRLOGAHLTSLSHEEQVFNVRV-----GHDYQ 2237
QY 76 WIGLRRREEKQSNSTACQDLVAWTGDSISQFNWVYDEP-----SCGSEVCVVMYHQSAP 131
DB 2238 WIGL-----NDKMFEDHFRWTDGSLQYENRPNQDSFSSAGEDCVIIWHENG-- 2287
QY 132 AGIGGYPYFMQNDRCNMKNFICKYS-----DEKPAVPSRAEAGE 172
DB 2288 -----QWNVPCNVHLTYTCKGTVACGQPPVVENAKTFGK 2323
RESULT 9
A47171
chondroitin sulfate proteoglycan PG-M core protein - chicken
C:Species: Gallus gallus (Chicken)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A47171
R:Shinomura, T.; Nishida, Y.; Ito, K.; Kimata, K.
J. Biol. Chem. 268, 14461-14469, 1993
A:Title: cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan expressed durin
A:Reference number: A47171; MUID:93300846; PMID:8314802
A:Accession: A47171
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-3562 <SHI>
A:Cross-references: GB:D13542; NID:g391643; PIDN:BAAO2742.1; PID:g391644
A:Experimental source: stage 22-23 developing limb buds
A:Note: sequence extracted from NCBI backbone (NCBIN:134456, NCBIP:134457)
C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lecti
F:166-243/Domain: link protein repeat homology <LNK1>
F:264-345/Domain: link protein repeat homology <LNK2>
F:3258-3289/Domain: EGF homology <EGF1>
F:3296-3327/Domain: EGF homology <EGF>
F:3334-3454/Domain: C-type lectin homology <LCH>
F:3461-3517/Domain: complement factor H repeat homology <FHD>
Query Match 15.3%; Score 171; DB 2; Length 3562;
Best Local Similarity 28.5%; Pred. No. 2.4e-06;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
QY 17 QRPCKVIYFHDTSRLNFEFEAKEACRRDGGQVLSIESEDEQKLEKFIENLLPSDGF- 75
DB 3342 QGQCYK--YF---AHRRTWDAARECRLOGAHLTSLSHEEQVFNRI-----GHDYQ 3389
QY 76 WIGLRRREEKQSNSTACQDLVAWTGDSISQFNWVYDEP-----SCGSEVCVVMYHQSAP 131
DB 3390 WIGL-----NDKMFEDHFRWTDGSLQYENRPNQDSFSSAGEDCVIIWHENG-- 3439
QY 132 AGIGGYPYFMQNDRCNMKNFICKYS-----DEKPAVPSRAEAGE 172
DB 3440 -----QWNVPCNVHLTYTCKGTVACGQPPVVENAKTFGK 3475
RESULT 10
T42710
mannose receptor, macrophage - mouse
N:Alternate names: lambda lectin; phospholipase A2 receptor
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000

C;Accession: T42710
R;Wu, K.; Yuan, J.; Lasky, L.A.
J. Biol. Chem. 271, 21323-21330, 1996
A;Title: Characterization of a novel member of the macrophage mannose receptor type C
A;Reference number: 222235; MUID:96355501; PMID:8702911
A;Accession: T42710
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1479 <WUK>
A;Cross-references: EMBL:U56734; NID:g1336073; PID:g1336074; PIDN:AAC52729.1
C;Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II re
C;Keywords: membrane protein; receptor
F;186-227/Domain: fibronectin type II repeat homology <2FR>

Query Match 15.3%; Score 170.5; DB 2; Length 1479;
Best Local Similarity 31.4%; Pred. No. 9.6e-07;
Matches 48; Conservative 22; Mismatches 48; Indels 35; Gaps 7;

Qy 7 SQPVCRGGTQPCYKVIYFHDTSRLNLFEEAKACRRDGGQLVSISEDEQKLIKPIE 66
Db 384 SWQPF-----QGHCYRL-----QAERKSQWESKRAKLRGGDLISIHSMAELEFTRKQIK 433
Qy 67 NLLPSDGDFTWGLRRREEKQSNSTACODLYAWTDGSIQFRNWWYVDEPS-----CGSEVCVV 123
Db 434 QEVE---ELWIGL-----NDLKQNNFWSQGLSVFTHHPFPBNFRDLSDCVT 482
Qy 124 MYHPSAPAGIGGYPMFQWDDRCNMKNFICK 156
Db 483 IW-----GPEG-----RWNDSPCNQSLPSICK 504

RESULT 11
A39808
proteoglycan core protein, cartilage - bovine (fragments)
N;Alternate names: aggrecan; aggregating cartilage proteoglycan
C;Species: Bos primigenius taurus (cattle)
C;Date: 20-Mar-1992 #sequence_revision 23-Mar-1995 #text_change 13-Aug-1999
C;Accession: A34234; A27752; A39808; A27751; E29164; B27751; C27751; E27751; F27
R;Antonsson, P.; Heinemann, D.; Oldberg, A.
J. Biol. Chem. 264, 16170-16173, 1989
A;Title: The keratan sulfate-enriched region of bovine cartilage proteoglycan consists
A;Reference number: A34234; MUID:89380219; PMID:2528543
A;Accession: A34234
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 128-621 <ANT>
A;Cross-references: GB:J05028
R;Oldberg, A.; Antonsson, P.; Heinemann, D.
Biochem. J. 243, 255-259, 1987
A;Title: The partial amino acid sequence of bovine cartilage proteoglycan, deduced from
A;Reference number: A27752; MUID:87270630; PMID:3111460
A;Accession: A27752
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 622-1340 <OLD>
R;Sandy, J.D.; Boynton, R.E.; Flannery, C.R.
J. Biol. Chem. 266, 8198-8205, 1991
A;Title: Analysis of the catabolism of aggrecan in cartilage explants by quantitation of
A;Reference number: A39808; MUID:91217051; PMID:2022637
A;Accession: A39808
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-28;59-82;131-137; 'QSET', 142-149; 196-207; 226-249; 1137-1143; 1252-1267; 1274-1
R;Perin, J.P.; Bonnet, P.; Jolles, P.
FEBS Lett. 206, 73-77, 1986
A;Title: Structural relationship between link proteins and proteoglycan monomers.
A;Reference number: A27751; MUID:87005253; PMID:3530809
A;Accession: A27751
A;Molecule type: protein
A;Residues: 29-58;74-130; 174-175; 'A', 177-204; 208-225 <PER>
R;Perin, J.P.; Bonnet, P.; Jolles, P.
FEBS Lett. 176, 37-42, 1984
A;Title: Sequence data concerning the protein core of the cartilage proteoglycan monomer

A;Reference number: A91327; MUID:85027710; PMID:6489519
A;Accession: E29164
A;Molecule type: protein
A;Residues: 1230-1249 <PE2>
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG
C;Keywords: glycoprotein
F;8-28/Domain: link protein repeat homology (fragment) <LNK1>
F;29-58/Domain: link protein repeat homology (fragment) <LNK2>
F;80-146/Domain: link protein repeat homology (fragments) <LNK3>
F;167-248/Domain: link protein repeat homology <LNK4>
F;1130-1250/Domain: C-type lectin homology <LCH>
F;1257-1313/Domain: complement factor H repeat homology <FHD>

Query Match 14.2%; Score 158.5; DB 2; Length 1340;
Best Local Similarity 26.8%; Pred. No. 9.9e-06;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;

Qy 9 QPVCRGG---TQPCYKVIYFHDTSRLNLFEEAKACRRDGGQLVSISEDEQKLIKPI 65
Db 1127 QKLCCEGWTKFQGHCYR--HFPD---RATWVDAESQCRKQKQSHLSSIVTPEEQ-----EFV 1177
Qy 66 ENLLPSDGDFTWGLRRREEKQSNSTACODLYAWTDGSIQFRNWWYVDEP-----SCGSEV 120
Db 1178 NN---NAQDYQWIGL-----NDKTIEGDFRWSGDGHSLSQFENWRPNQDPNFFATGDC 1226
Qy 121 CVVMYHPSAPAGIGGYPMFQWDDRCNMKNFICKYS---DEKPAVPSREAGEETE 175
Db 1227 VVMIWHEKG-----EWNDVPCNYQLPFTCKGTACGPPVVEHARIFGQKKD 1274

RESULT 12
T42630
aggrecan - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C;Accession: T42630
R;Hering, T.M.; Kollar, J.; Huynh, T.D.
submitted to the EMBL Data Library, September 1996
A;Description: Complete coding sequence of bovine aggrecan: comparative structural analy
A;Reference number: 222182
A;Accession: T42630
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-2327 <HER>
A;Cross-references: EMBL:U76615; NID:g1730259; PID:g1730260; PIDN:AAB38524.1
A;Experimental source: articular chondrocytes
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG
C;Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glycoprot

Query Match 14.2%; Score 158.5; DB 2; Length 2327;
Best Local Similarity 26.8%; Pred. No. 1.9e-05;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;

Qy 9 QPVCRGG---TQPCYKVIYFHDTSRLNLFEEAKACRRDGGQLVSISEDEQKLIKPI 65
Db 2114 QKLCCEGWTKFQGHCYR--HFPD---RATWVDAESQCRKQKQSHLSSIVTPEEQ-----EFV 2164
Qy 66 ENLLPSDGDFTWGLRRREEKQSNSTACODLYAWTDGSIQFRNWWYVDEP-----SCGSEV 120
Db 2165 NN---NAQDYQWIGL-----NDKTIEGDFRWSGDGHSLSQFENWRPNQDPNFFATGDC 2213
Qy 121 CVVMYHPSAPAGIGGYPMFQWDDRCNMKNFICKYS---DEKPAVPSREAGEETE 175
Db 2214 VVMIWHEKG-----EWNDVPCNYQLPFTCKGTACGPPVVEHARIFGQKKD 2261

RESULT 13
A39086
aggrecan precursor, cartilage long splice form [validated] - human
N;Alternate names: chondroitin sulfate proteoglycan 1; large aggregating proteoglycan; p
N;Contains: aggrecan cartilage short splice form
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000

Search completed: December 22, 2003, 16:14:42
Job time : 11.0151 secs

RESULT 15
A28452
proteoglycan core protein precursor, cartilage - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1989 #sequence,revision 30-Jun-1989 #text_change 21-Jan-2000
C;Accession: A92623; A23835; A28453; A28095; A28452
R;Doegge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.
J. Biol. Chem. 262, 17757-17767, 1987
A;Title: Complete primary structure of the rat cartilage proteoglycan core protein deduced from complementary DNA
A;Reference number: A92623; MUID:88087070; PMID:3693370
A;Accession: A92623
A;Molecule type: mRNA
A;Residues: 1-2124 <DOE>
R;Doegge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.
J. Biol. Chem. 263, 10040a, 1988
A;Reference number: A30069
A;Contents: annotation; revision to residue 698
R;Doegge, K.; Fernandez, P.; Hassell, J.R.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 261, 8108-8111, 1986
A;Title: Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan core protein deduced from complementary DNA
A;Reference number: A23835; MUID:86250698; PMID:2424893
A;Accession: A23835
A;Molecule type: mRNA
A;Residues: 1856-2124 <DO2>
A;Cross-references: GB:M13518; NID:G206104; PIDN:AAA41836.1; PID:G206105
R;Neame, P.J.; Christner, J.E.; Baker, J.R.
J. Biol. Chem. 262, 17768-17778, 1987
A;Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan amino-terminal domain
A;Reference number: A28453; MUID:88087071; PMID:3693371
A;Accession: A28453
A;Molecule type: protein
A;Residues: 20-37,'W',39-60,'E',62-64,'X',66-69;70-83;84,89-148,'L',150-238,'S',240,'A',241-243
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EC:3.2.1.160
C;Keywords: glycoprotein
F;1-19/Domain: signal sequence
F;20-2124/Product: proteoglycan core protein #status predicted <SIG>
F;44-135/Domain: immunoglobulin homology <IMM>
F;170-247/Domain: link protein repeat homology <LNK1>
F;268-349/Domain: link protein repeat homology <LNK2>
F;504-581/Domain: link protein repeat homology <LNK3>
F;602-683/Domain: link protein repeat homology <LNK4>
F;1914-2034/Domain: C-type lectin homology <LCH>
F;2041-2097/Domain: complement factor H repeat homology <FHD>

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:07:10 ; Search time 19.6848 Seconds
(without alignments)
1954.412 Million cell updates/sec

Title: US-09-887-855-5

Perfect score: 1115

Sequence: 1 ATGRLLSGQPVCVCGGTQPC.....EEDAKTKFKESREALNLAY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------------------|-------------------|
| 1 | 1115 | 100.0 | 206 | 9 US-09-887-855-5 | Sequence 5, Appli |
| 2 | 1115 | 100.0 | 374 | 9 US-09-887-855-2 | Sequence 2, Appli |
| 3 | 1115 | 100.0 | 374 | 12 US-10-094-749-2090 | Sequence 2090, Ap |
| 4 | 1115 | 100.0 | 374 | 15 US-10-149-819-15 | Sequence 15, Appl |
| 5 | 1111 | 99.6 | 374 | 12 US-10-094-749-2142 | Sequence 2142, Ap |
| 6 | 1101 | 98.7 | 382 | 10 US-09-909-320-137 | Sequence 137, App |
| 7 | 1101 | 98.7 | 382 | 10 US-09-909-088B-137 | Sequence 137, App |
| 8 | 1101 | 98.7 | 382 | 10 US-09-905-291A-137 | Sequence 137, App |
| 9 | 1101 | 98.7 | 382 | 10 US-09-902-853-137 | Sequence 137, App |
| 10 | 1101 | 98.7 | 382 | 10 US-09-907-824-137 | Sequence 137, App |
| 11 | 1101 | 98.7 | 382 | 10 US-09-907-841-137 | Sequence 137, App |
| 12 | 1101 | 98.7 | 382 | 11 US-09-904-011-137 | Sequence 137, App |
| 13 | 1101 | 98.7 | 382 | 11 US-09-906-742-137 | Sequence 137, App |
| 14 | 1101 | 98.7 | 382 | 11 US-09-906-838-137 | Sequence 137, App |
| 15 | 1101 | 98.7 | 382 | 11 US-09-907-613-137 | Sequence 137, App |

| | | | | | |
|----|------|------|-----|-----------------------|-------------------|
| 16 | 1101 | 98.7 | 382 | 11 US-09-907-942-137 | Sequence 137, App |
| 17 | 1101 | 98.7 | 382 | 11 US-09-904-859-137 | Sequence 137, App |
| 18 | 1101 | 98.7 | 382 | 11 US-09-909-204-137 | Sequence 137, App |
| 19 | 1101 | 98.7 | 382 | 11 US-09-904-820-137 | Sequence 137, App |
| 20 | 1101 | 98.7 | 382 | 11 US-09-904-786-137 | Sequence 137, App |
| 21 | 1101 | 98.7 | 382 | 11 US-09-906-646-137 | Sequence 137, App |
| 22 | 1101 | 98.7 | 382 | 11 US-09-906-700-137 | Sequence 137, App |
| 23 | 1101 | 98.7 | 382 | 11 US-09-903-786-137 | Sequence 137, App |
| 24 | 1101 | 98.7 | 382 | 11 US-09-903-903-137 | Sequence 137, App |
| 25 | 1101 | 98.7 | 382 | 11 US-09-903-749A-137 | Sequence 137, App |
| 26 | 1101 | 98.7 | 382 | 11 US-09-904-119-137 | Sequence 137, App |
| 27 | 1101 | 98.7 | 382 | 11 US-09-904-956-137 | Sequence 137, App |
| 28 | 1101 | 98.7 | 382 | 11 US-09-903-736-137 | Sequence 137, App |
| 29 | 1101 | 98.7 | 382 | 11 US-09-907-794-137 | Sequence 137, App |
| 30 | 1101 | 98.7 | 382 | 11 US-09-903-943-137 | Sequence 137, App |
| 31 | 1101 | 98.7 | 382 | 11 US-09-904-462-137 | Sequence 137, App |
| 32 | 1101 | 98.7 | 382 | 11 US-09-907-925-137 | Sequence 137, App |
| 33 | 1101 | 98.7 | 382 | 11 US-09-902-692-137 | Sequence 137, App |
| 34 | 1101 | 98.7 | 382 | 11 US-09-903-520-137 | Sequence 137, App |
| 35 | 1101 | 98.7 | 382 | 11 US-09-905-056-137 | Sequence 137, App |
| 36 | 1101 | 98.7 | 382 | 11 US-09-909-064-137 | Sequence 137, App |
| 37 | 1101 | 98.7 | 382 | 11 US-09-904-553-137 | Sequence 137, App |
| 38 | 1101 | 98.7 | 382 | 11 US-09-905-381-137 | Sequence 137, App |
| 39 | 1101 | 98.7 | 382 | 11 US-09-905-088-137 | Sequence 137, App |
| 40 | 1101 | 98.7 | 382 | 11 US-09-907-575-137 | Sequence 137, App |
| 41 | 1101 | 98.7 | 382 | 11 US-09-905-075-137 | Sequence 137, App |
| 42 | 1101 | 98.7 | 382 | 11 US-09-902-759-137 | Sequence 137, App |
| 43 | 1101 | 98.7 | 382 | 11 US-09-902-634-137 | Sequence 137, App |
| 44 | 1101 | 98.7 | 382 | 11 US-09-902-713-137 | Sequence 137, App |
| 45 | 1101 | 98.7 | 382 | 11 US-09-907-979-137 | Sequence 137, App |

ALIGNMENTS

RESULT 1
US-09-887-855-5
; Sequence 5, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-887-855-5

Query Match 100.0%; Score 1115; DB 9; Length 206;
Best Local Similarity 100.0%; Pred. No. 7, 7e-105;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----------------------|
| QY | 1 | ATGRLLSGQPVCVCGGTQPCYKVIYFHDTSRLNFEAEACRRDGGQLVSI | EDEQKL 60 |
| Db | 1 | ATGRLLSGQPVCVCGGTQPCYKVIYFHDTSRLNFEAEACRRDGGQLVSI | EDEQKL 60 |
| QY | 61 | IEKFTIENLLPSDGFHWIGLRRREEKOSNSTACODLYAWTDGSI | SQFRNNYVDEPSCGSEV 120 |
| Db | 61 | IEKFTIENLLPSDGFHWIGLRRREEKOSNSTACODLYAWTDGSI | SQFRNNYVDEPSCGSEV 120 |
| QY | 121 | CVVMYHQFSAPAGIGPYMFQWDDRCNMKNNFICKYDEKPAVPSRE | AGEETEELTPV 180 |
| Db | 121 | CVVMYHQFSAPAGIGPYMFQWDDRCNMKNNFICKYDEKPAVPSRE | AGEETEELTPV 180 |
| QY | 181 | LPEETQEDAKTKFKESREALNLAY 206 | |
| Db | 181 | LPEETQEDAKTKFKESREALNLAY 206 | |

```
RESULT 2
US-09-887-855-2
; Sequence 2, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-887-855-2
Query Match      100.0%; Score 1115; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.7e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEFEAKEACRRDGGOLVSISEDEQKL 60
Db 22 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEFEAKEACRRDGGOLVSISEDEQKL 81

QY 61 IEKFIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIISQFRNWWYDEPSCGSEV 120
Db 82 IEKFIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIISQFRNWWYDEPSCGSEV 141

QY 121 CVVMYHQPAPAGIGGPPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPV 180
Db 142 CVVMYHQPAPAGIGGPPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPV 201

QY 181 LPEETQEDAKKTFKESREAAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAAALNLAY 227

RESULT 3
US-10-094-749-2090
; Sequence 2090, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KENJI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
```

```
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2090
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2090
Query Match      100.0%; Score 1115; DB 12; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.7e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEFEAKEACRRDGGOLVSISEDEQKL 60
Db 22 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEFEAKEACRRDGGOLVSISEDEQKL 81

QY 61 IEKFIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIISQFRNWWYDEPSCGSEV 120
Db 82 IEKFIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIISQFRNWWYDEPSCGSEV 141

QY 121 CVVMYHQPAPAGIGGPPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPV 180
Db 142 CVVMYHQPAPAGIGGPPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPV 201

QY 181 LPEETQEDAKKTFKESREAAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAAALNLAY 227

RESULT 4
US-10-149-819-15
; Sequence 15, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dyoung Aina M.
; APPLICANT: SHAH, Purvi
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PF-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; FEATURE:
; OTHER INFORMATION: Incyte ID No. US20030044913A1 3143411CD1
US-10-149-819-15
Query Match      100.0%; Score 1115; DB 15; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.7e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEFEAKEACRRDGGOLVSISEDEQKL 60
Db 22 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEFEAKEACRRDGGOLVSISEDEQKL 81

QY 61 IEKFIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIISQFRNWWYDEPSCGSEV 120
Db 82 IEKFIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIISQFRNWWYDEPSCGSEV 141
```

Oy 121 CVVMYHQPAPAGIGGYPYFQWDDRCNMKNFICKYSDKPAVPSREAGEETLTPV 180
Db 142 CVVMYHQPAPAGIGGYPYFQWDDRCNMKNFICKYSDKPAVPSREAGEETLTPV 201
Oy 181 LPEETOEDAKTKFKESREAAALNLAY 206
Db 202 LPEETOEDAKTKFKESREAAALNLAY 227

RESULT 5
US-10-094-749-2142
; Sequence 2142, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KETICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR FILING DATE: 2002-01-24
; PRIOR FILING DATE: 2002-01-24
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2142
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2142

Query Match 99.6%; Score 1111; DB 12; Length 374;
Best Local Similarity 99.5%; Pred. No. 4.4e-104;
Matches 205; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 ATGRLSGQPCVCGGTORPCYKVLYFHDTSRRLNFEAKACRRDGGQLVSISEDEQKL 60
Db 22 ATGRLSGQPCVCGGTORPCYKVLYFHDTSRRLNFEAKACRRDGGQLVSISEDEQKL 81
Oy 61 IERFIENLLPSDGFDFWGLRRREKQSNSTACQDLVYAWTDGSIQFRNWWYVDEPSCGSEV 120
Db 82 IERFIENLLPSDGFDFWGLRRREKQSNSTACQDLVYAWTDGSIQFRNWWYVDEPSCGSEV 141
Oy 121 CVVMYHQPAPAGIGGYPYFQWDDRCNMKNFICKYSDKPAVPSREAGEETLTPV 180
Db 142 CVVMYHQPAPAGIGGYPYFQWDDRCNMKNFICKYSDKPAVPSREAGEETLTPV 201
Oy 181 LPEETOEDAKTKFKESREAAALNLAY 206
Db 202 LPEETOEDAKTKFKESREAAALNLAY 227

RESULT 6
US-09-909-320-137
; Sequence 137, Application US/09909320

Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-320-137


```

, PRIOR FILING DATE: 1999-10-05
, PRIOR APPLICATION NUMBER: PCT/US99/28214
, PRIOR FILING DATE: 1999-11-29
, PRIOR APPLICATION NUMBER: PCT/US99/28313
, PRIOR FILING DATE: 1999-11-30
, PRIOR APPLICATION NUMBER: PCT/US99/28564
, PRIOR FILING DATE: 1999-12-02
, PRIOR APPLICATION NUMBER: PCT/US99/28565
, PRIOR FILING DATE: 1999-12-02
, PRIOR APPLICATION NUMBER: PCT/US99/30095
, PRIOR FILING DATE: 1999-12-16
, PRIOR APPLICATION NUMBER: PCT/US99/30911
, PRIOR FILING DATE: 1999-12-20
, PRIOR APPLICATION NUMBER: PCT/US99/30999
, PRIOR FILING DATE: 1999-12-20
, PRIOR APPLICATION NUMBER: PCT/US00/00219
, PRIOR FILING DATE: 2000-01-05
, NUMBER OF SEQ ID NOS: 423
, SEQ ID NO 137
, LENGTH: 382
, TYPE: PRT
, ORGANISM: Homo sapiens
US-09-909-088B-137

```

| Query Match | 98.7% | Score 1101 | DB 10 | Length 382 | |
|-----------------------|--|--|--|------------|-----|
| Best Local Similarity | 96.3% | Pred. No. 4.6e-103 | | | |
| Matches 206 | Conservative 0 | Mismatches 0 | Indels 8 | Gaps 1 | |
| Qy 1 | ATGRLLS----- | GOPVCRGGTQRCYKVIY | FHDTSSRRLNFEAKACRRDGGQLVSI | 52 | |
| | | | | | |
| Db | 22 | ATGELLSASDLDRGGOPVCEGGTQRCYKVIY | FHDTSSRRLNFEAKACRRDGGQLVSI | 81 | |
| | | | | | |
| Qy 53 | ESEDEQKLI | EKFIEIENLPSDGFWIGLRRREEKOSNSTACQDLYAWTDGSI | SQFRNYYWD | 112 | |
| | | | | | |
| Db | 82 | ESEDEQKLI | EKFIEIENLPSDGFWIGLRRREEKOSNSTACQDLYAWTDGSI | SQFRNYYWD | 141 |
| | | | | | |
| Qy 113 | EPSCGSEVCVMYHQPASAPAGIGPGYMFQWNNDRCKMKNFICKYSDEKPAVPSREAGE | 172 | | | |
| | | | | | |
| Db | 142 | EPSCGSEVCVMYHQPASAPAGIGPGYMFQWNNDRCKMKNFICKYSDEKPAVPSREAGE | 201 | | |
| | | | | | |
| Qy 173 | ETELTTPVLPPEETOEDAKKTFKESREAAALNLAY | 206 | | | |
| | | | | | |
| Db | 202 | ETELTTPVLPPEETOEDAKKTFKESREAAALNLAY | 235 | | |
| | | | | | |

RESULT 8
US-09-905-219A-137
Sequence 137, Application US/09905291A
Patent No. US20030160374A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Kenneth L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel

```
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-905-291A-137

Query Match      98.7%; Score 1101; DB 10; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQPVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSI 52
Db 22 ATGRLLSASLDLRGGQPVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSI 81
Qy 53 ESEDEQKLIKFNIENLLPSDGFVIGLRRRREKQSNSTACQDLVWYTDGSIQFRNWTVD 112
Db 82 ESEDEQKLIKFNIENLLPSDGFVIGLRRRREKQSNSTACQDLVWYTDGSIQFRNWTVD 141
Qy 113 EPSCGSEVVMYHQPSPAGIGGYPWFQWDDRCNKNKFNICKYSDEKPAVPSREAGE 172
Db 142 EPSCGSEVVMYHQPSPAGIGGYPWFQWDDRCNKNKFNICKYSDEKPAVPSREAGE 201
Qy 173 ETELTTPVLPRETQEDAKKTFKESREAAALNLAY 206
Db 202 ETELTTPVLPRETQEDAKKTFKESREAAALNLAY 235

RESULT 9
US-09-902-853-137
; Sequence 137, Application US/09902853
; Publication No. US20020192659A1

; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,853
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US/09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-902-853-137

Query Match      98.7%; Score 1101; DB 10; Length 382;
```

```
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEFEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLDRGGQVCRGGTQPCYKVIYFHDTSRRLNFEFEAKEACRRDGGQLVSI 81
QY 53 ESEDEOKLIEKFENLLPSDGFWIGLRRREEKQSNSTACQDLYAWMTDGSISQFRNMYVD 112
Db 82 ESEDEOKLIEKFENLLPSDGFWIGLRRREEKQSNSTACQDLYAWMTDGSISQFRNMYVD 141
QY 113 EPCSGSEVCVMYHQSPAGIGPPYMFQWDDRCNMKNFICKYSDEKPAVPSRAEGE 172
Db 142 EPCSGSEVCVMYHQSPAGIGPPYMFQWDDRCNMKNFICKYSDEKPAVPSRAEGE 201
QY 173 ETELTPVLPEETOEDAKKTFKESREAAALNLAY 206
Db 202 ETELTPVLPEETOEDAKKTFKESREAAALNLAY 235

RESULT 10
US-09-907-824-137
; Sequence 137, Application US/09907824
; Publication No. US20020197671A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,824
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
```

```
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO. 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-907-824-137

Query Match 98.7%; Score 1101; DB 10; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEFEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLDRGGQVCRGGTQPCYKVIYFHDTSRRLNFEFEAKEACRRDGGQLVSI 81
QY 53 ESEDEOKLIEKFENLLPSDGFWIGLRRREEKQSNSTACQDLYAWMTDGSISQFRNMYVD 112
Db 82 ESEDEOKLIEKFENLLPSDGFWIGLRRREEKQSNSTACQDLYAWMTDGSISQFRNMYVD 141
QY 113 EPCSGSEVCVMYHQSPAGIGPPYMFQWDDRCNMKNFICKYSDEKPAVPSRAEGE 172
Db 142 EPCSGSEVCVMYHQSPAGIGPPYMFQWDDRCNMKNFICKYSDEKPAVPSRAEGE 201
QY 173 ETELTPVLPEETOEDAKKTFKESREAAALNLAY 206
Db 202 ETELTPVLPEETOEDAKKTFKESREAAALNLAY 235

RESULT 11
US-09-907-841-137
; Sequence 137, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
```

```
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,841
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-907-841-137

Query Match          98.7%; Score 1101; DB 10; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGOLVSI 52
    |||||
Db 22 ATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGOLVSI 81

QY 53 ESEDEQKLIETFIENLLPSDGFHWIGLRREEKQSNSTACQDLYAWTDGSIQFRNYYVD 112
    |||||
Db 82 ESEDEQKLIETFIENLLPSDGFHWIGLRREEKQSNSTACQDLYAWTDGSIQFRNYYVD 141

QY 113 EPCSCSEVCVMYHQPAPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAGE 172
    |||||
Db 142 EPCSCSEVCVMYHQPAPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAGE 201

QY 173 ETELTPVLPEETQEDAKKTFKESREAAALNLAY 206
    |||||
Db 202 ETELTPVLPEETQEDAKKTFKESREAAALNLAY 235

RESULT 12
US-09-904-011-137
; Sequence 137, Application US/05904011
; Publication No. US2003000330A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
```

```
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mathier, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
; ORGANISM: Homo Sapien
US-09-904-011-137

Query Match          98.7%; Score 1101; DB 11; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGOLVSI 52
    |||||
Db 22 ATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGOLVSI 81

QY 53 ESEDEQKLIETFIENLLPSDGFHWIGLRREEKQSNSTACQDLYAWTDGSIQFRNYYVD 112
    |||||
Db 82 ESEDEQKLIETFIENLLPSDGFHWIGLRREEKQSNSTACQDLYAWTDGSIQFRNYYVD 141
```


; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapiens
; US-09-906-838-137

Query Match 98.7%; Score 1101; DB 11; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSI 52
DB 22 ATGRLLSASDLRLGGQPCVCRGGTQPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSI 81
QY 53 ESEDEQKLIKFIEHLPLPSDGFWIGLRRREEKOSNSTACQDLVYAWTDGSIQPRNNYVD 112
DB 82 ESEDEQKLIKFIEHLPLPSDGFWIGLRRREEKOSNSTACQDLVYAWTDGSIQPRNNYVD 141
QY 113 EPSCGSEVCMVYHOPAPAGIGGYPYFQWDDRCNNKNNFICKYSDEKPAVPSREAGE 172
DB 142 EPSCGSEVCMVYHOPAPAGIGGYPYFQWDDRCNNKNNFICKYSDEKPAVPSREAGE 201
QY 173 ETELTTPVLPETQEDAKTKFESREAAALNLAY 206
DB 202 ETELTTPVLPETQEDAKTKFESREAAALNLAY 235

RESULT 15

US-09-907-613-137
; Sequence 137, Application US/09907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Akenkazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone

; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,613
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-907-613-137

Query Match 98.7%; Score 1101; DB 11; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSI 52
DB 22 ATGRLLSASDLRLGGQPCVCRGGTQPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSI 81

| | | | | | | | | | | | | | |
|----|-----|-----------|-------|--------|-------|---------|--------|---------|--------|--------|--------|---------|-----|
| Qy | 53 | ESEDEQKLI | EKFIE | NIENLL | PSDGD | FWIGL | RRREEK | QSNSTAC | QDL | YAWTD | GSISQ | FRNWYVD | 112 |
| Db | 82 | ESEDEQKLI | EKFIE | NIENLL | PSDGD | FWIGL | RRREEK | QSNSTAC | QDL | YAWTD | GSISQ | FRNWYVD | 141 |
| Qy | 113 | EPSCGSE | VCVVM | YHOPS | APAGI | GGPYM | FQWDD | RCNMKN | NFICKY | SDEKPA | VPSPRE | AEGE | 172 |
| Db | 142 | EPSCGSE | VCVVM | YHOPS | APAGI | GGPYM | FQWDD | RCNMKN | NFICKY | SDEKPA | VPSPRE | AEGE | 201 |
| Qy | 173 | ETELTPVL | PEETO | EDAKT | FKESR | EALNLAY | 206 | | | | | | |
| Db | 202 | ETELTPVL | PEETO | EDAKT | FKESR | EALNLAY | 235 | | | | | | |

Search completed: December 22, 2003, 16:16:49
 Job time : 20.6848 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:05:09 ; Search time 6.46605 Seconds
(without alignments)
1476.391 Million cell updates/sec

Title: US-09-887-855-2_COPY_25_227

Perfect score: 1100

Sequence: 1 RLLSQPVCGRGTRPCYKV.....BEDAKTKFKESREAAALNAY 203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 574 | 52.2 | 273 | 1 | CHOD MOUSE |
| 2 | 561.5 | 51.0 | 273 | 1 | CHOD HUMAN |
| 3 | 182 | 16.5 | 1456 | 1 | MANR_HUMAN |
| 4 | 177 | 16.1 | 1268 | 1 | PGCN MOUSE |
| 5 | 176 | 16.0 | 1321 | 1 | PGCN HUMAN |
| 6 | 174.5 | 15.9 | 3381 | 1 | PGCV BOVIN |
| 7 | 174 | 15.8 | 1257 | 1 | PGCV RAT |
| 8 | 174 | 15.8 | 2728 | 1 | PGCV RAT |
| 9 | 174 | 15.8 | 3358 | 1 | PGCV MOUSE |
| 10 | 174 | 15.8 | 3396 | 1 | PGCV HUMAN |
| 11 | 171 | 15.5 | 3562 | 1 | PGCV CHICK |
| 12 | 164 | 14.9 | 643 | 1 | CD93 RAT |
| 13 | 158.5 | 14.4 | 2364 | 1 | PGCA BOVIN |
| 14 | 158.5 | 14.4 | 2415 | 1 | PGCA HUMAN |
| 15 | 155.5 | 14.1 | 2333 | 1 | PGCA CANFA |
| 16 | 154.5 | 14.0 | 612 | 1 | LEW2 MOUSE |
| 17 | 153.5 | 14.0 | 644 | 1 | CD93 MOUSE |
| 18 | 153.5 | 14.0 | 2124 | 1 | PGCA RAT |
| 19 | 152 | 13.8 | 912 | 1 | PGCB BOVIN |
| 20 | 151.5 | 13.8 | 652 | 1 | CD93 HUMAN |
| 21 | 151 | 13.7 | 197 | 1 | CLF1_HUMAN |
| 22 | 149 | 13.5 | 321 | 1 | FCB2_HUMAN |
| 23 | 147.5 | 13.4 | 2132 | 1 | PGCA MOUSE |
| 24 | 147 | 13.4 | 883 | 1 | PGCB MOUSE |
| 25 | 146 | 13.3 | 2109 | 1 | PGCA CHICK |
| 26 | 145 | 13.2 | 158 | 1 | LEGC TRIST |
| 27 | 145 | 13.2 | 883 | 1 | PGCB RAT |
| 28 | 144.5 | 13.1 | 173 | 1 | LEC2_MEGRO |
| 29 | 144.5 | 13.1 | 372 | 1 | LEMI RAT |
| 30 | 143.5 | 12.9 | 372 | 1 | LEMI MOUSE |
| 31 | 141.5 | 12.9 | 331 | 1 | FCB2 MOUSE |
| 32 | 141.5 | 12.9 | 549 | 1 | LEW2 RAT |
| 33 | 140.5 | 12.8 | 162 | 1 | LEC3_MEGRO |
| | | | | | P07439 megabalanus |

| | | | | | |
|----|-------|------|-----|---|------------|
| 34 | 140 | 12.7 | 370 | 1 | LEMI BOVIN |
| 35 | 139 | 12.6 | 248 | 1 | PSPA HUMAN |
| 36 | 139 | 12.6 | 283 | 1 | LECA_SARPE |
| 37 | 138.5 | 12.6 | 152 | 1 | IXA_TRIFL |
| 38 | 138.5 | 12.6 | 202 | 1 | TETN MOUSE |
| 39 | 137.5 | 12.5 | 372 | 1 | LEMI MACMU |
| 40 | 137.5 | 12.5 | 372 | 1 | LEMI PACHU |
| 41 | 136 | 12.4 | 175 | 1 | LITH BOVIN |
| 42 | 135 | 12.3 | 165 | 1 | LIT1_MOUSE |
| 43 | 134.5 | 12.2 | 132 | 1 | ACAL_ANGAN |
| 44 | 134.5 | 12.2 | 372 | 1 | LEMI_PONPY |
| 45 | 134.5 | 12.2 | 485 | 1 | LEM2_BOVIN |

ALIGNMENTS

RESULT 1

CHOD_MOUSE

ID CHOD_MOUSE STANDARD; PRT; 273 AA.
AC Q9CXM0: Q8VI31;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chondlectin precursor (Transmembrane protein MT75).
GN CHODL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Weng L., Smith P., Hubner R., Wouters J., Merregaert J.;
RT "Mt75, a low expressed c-type lectin gene involving in
chondrogenesis";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Niehi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saico T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,
RA Brownstein M.J., Bult C., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaeerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyono-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
CC - SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC - SIMILARITY: Contains 1 C-type lectin family domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF311699; AAL50354.1; -


```

DR EMBL: AK014255; BAB29226.1; -.
DR HSP: P22897; 1EGG.
DR MGI: 2179069; Chodl.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE LECTIN 1; FALSE_NEG.
DR PROSITE: PS00041; C-TYPE LECTIN 2; 1.
KW Lectin; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 273 CHONDROLECTIN.
FT DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 217 237 POTENTIAL.
FT DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 179 C-TYPE LECTIN.
FT CARBOHYD 86 86 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 24 24 V -> W (IN REF. 2).
FT CONFLICT 179 179 T -> K (IN REF. 2).
SQ SEQUENCE 273 AA; E052D933F244F4C7 CRC64;

Query Match 52.2%; Score 574; DB 1; Length 273;
Best Local Similarity 57.4%; Pred. No. 4.1e-44;
Matches 109; Conservative 26; Mismatches 41; Indels 14; Gaps 5;

QY 1 RLLSGQPCVCGGTQPCYKVIYFHTSRRRLNFEAEACRQGGQGVLSIESDEOKLIEK 60
Db ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
23 RVVSGQKVFADYKPCYKWAYFHELSRVSFQEARLACESEGGVLLSLENAEQKLIES 82
QY 61 FIENLLP-----SDGFWTGLRRREKQNSACQDLYAWTQSGISQFRNVTVDPPSCGS 115
Db ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
83 MLQNLTKPGTGISDGDFTWGLLRSGDGT-SGACPDLYQWSDGSSQFRNVTVDPPSCGS 141
QY 116 EYCVVMYHOPSPAGTGGYPMQWDDRCNMKNFICKYSDE-KPAVPREAGSETELT 174
Db ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
142 EYCVVMYHOPSPAGTGGYPMQWDDRCNMKNFICKYSDE-KPAVPREAGSETELT 174
QY 175 TPVLPETQOE 184
Db ||| |||
197 NQ--FEETHE 204

RESULT 2
CHOD_HUMAN STANDARD; PRT; 273 AA.
AC Q9H9P2; Q9HCY3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Chondrolectin precursor (Transmembrane protein MT75) (PRED12
protein).
GN CHODL OR C21ORF68.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=22074930; PubMed=12079284;
RA Weng L., Smits P., Wauters J., Merregaert J.;
RT "Molecular cloning and characterization of human chondrolectin, a
novel type I transmembrane protein homologous to C-type lectins.";
RL Genomics 80:62-70(2002).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagayama K., Mitsuoyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordseik G., Hornischer K., Brandt P.,

```

```

RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Leirach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altachul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahay J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[4]
RP SEQUENCE OF 27-273 FROM N.A.
RA Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Takatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Watanabe S., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN TISSUE SPECIFICITY.
RX MEDLINE=21564202; PubMed=11707072;
RA Raymond A., Friedl M., Neergard Henriksen C., Chapot F., Deutsch S.,
RA Ucla C., Rossier C., Lytle R., Guipponi M., Antonarakis S.E.;
RT "From PREDs and open reading frames to cDNA isolation: revisiting the
human chromosome 21 transcription map.";
RL Genomics 78:46-54(2001).
CC -|- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -|- TISSUE SPECIFICITY: Found in spleen, testis, prostate and fetal
liver. Expression limited to vascular muscle of testis, smooth
muscle of prostate stroma, heart muscle, skeletal muscle, crypts
of small intestine, and red pulp of spleen.
CC -|- PTM: N-glycosylated.
CC -|- SIMILARITY: Contains 1 C-type lectin family domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: AF257472; AAL05981.1; -.
DR EMBL: AL163217; CAB90388.1; -.
DR EMBL: BC009418; AAB09418.1; -.
DR EMBL: AK022689; BAB14181.1; ALT_INIT.
DR HSSP: P22897; 1EGG.
DR Genew: HGNC:17807; CHODL.
DR MIM: 607247; -.
DR InterPro: IPR001304; Lectin_C.

```

```

DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; FALSE_NEG.
DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.
DR LECTIN; Transmembrane; Glycoprotein; Signal.
DR SIGNAL 1 21 POTENTIAL.
DR CHAIN 22 273 CHONDROLECTIN.
DR DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
DR TRANSMEM 217 237 POTENTIAL.
DR DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
DR DOMAIN 35 179 C-TYPE LECTIN.
DR CARBOHYD 86 86 N-LINKED (GLCNAC...) (POTENTIAL).
DR SEQUENCE 273 AA; 30431 MW; F4890AARF572A311 CRC64;

Query Match 51.0%; Score 561.5; DB 1; Length 273;
Best Local Similarity 60.1%; Pred. No. 5.4e-43;
Matches 101; Conservative 25; Mismatches 35; Indels 7; Gaps 3;

QY 1 RLISGQVCRGGTQPCYKVIYFHTSRRLNFEEAKACRRDGGQLVISEDQKLIK 60
DB 23 RVVSGQKVCFAFKPCYKWAYFHELSRVSFQEARLACESEGGVLLSLENEAQKLIK 82

QY 61 FIENLLP-----SGDGFWIGLRREKQSNSTACDLYAWTDGSIQSFNRYWYDPSGCS 115
DB 83 MLQNLTKEPGTGISDGDWFWIGLRNGDQGT-SGACPDLYQWSDGNSQYRWYTDPSGCS 141

QY 116 EVCVMYHQSPAGIGGYPWFQNDRCNMKNFNICKYSDE-KPAVP 162
DB 142 EKCVMYHQPTANPLGGPLYQWDDRCNMKNHYICKYPEINPTAP 189

RESULT 3
MANR_HUMAN
ID MANR_HUMAN STANDARD; PRT; 1456 AA.
AC P22897;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Macrophage mannose receptor precursor (MMR) (CD206 antigen).
GN MRC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Placenta;
RX MEDLINE=90324192; PubMed=2373685;
RA Taylor M.E., Conary J.T., Lennartz M.R., Stahl P.D., Drickamer K.;
RT "Primary structure of the mannose receptor contains multiple motifs
RT resembling carbohydrate-recognition domains.";
RL J. Biol. Chem. 265:12156-12162(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93052405; PubMed=1294118;
RA Kim S.J., Ruiz N., Bezouska K., Drickamer K.;
RT "Organization of the gene encoding the human macrophage mannose
RT receptor (MRC1).";
RL Genomics 14:721-727(1992).
RN [3]
RP STUDIES ON THE BINDING OF INDIVIDUAL LECTIN DOMAINS.
RX MEDLINE=92112893; PubMed=1730714;
RA Taylor M.E., Bezouska K., Drickamer K.;
RT "Contribution to ligand binding by multiple carbohydrate-recognition
RT domains in the macrophage mannose receptor.";
RL J. Biol. Chem. 267:1719-1726(1992).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 642-788.
RX MEDLINE=20347275; PubMed=10779515;
RA Feinberg H., Park-Snyder S., Kolatkar A.R., Heise C.T., Taylor M.E.,
RA Weis W.I.;
RT "Structure of a C-type carbohydrate recognition domain from the
RT macrophage mannose receptor.";

```

```

RL J. Biol. Chem. 275:21539-21548(2000).
CC -1- FUNCTION: MEDIATES THE ENDOCYTOSIS OF GLYCOPROTEINS BY
CC MACROPHAGES. IN SEVERAL RECOGNITION AND UPTAKE PROCESSES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: CRDS 1-3 HAVE AT MOST VERY WEAK AFFINITY FOR
CC CARBOHYDRATE. CRD 4 SHOWS THE HIGHEST AFFINITY BINDING AND HAS
CC MULTISPECIFICITY FOR A VARIETY OF MONOSACCHARIDES. AT LEAST 3 CRDS
CC (4, 5, AND 7) ARE REQUIRED FOR HIGH AFFINITY BINDING AND
CC ENDOCYTOSIS OF MULTIVALENT GLYCOCONJUGATES.
CC -1- SIMILARITY: Contains 8 C-type lectin family domains.
CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -1- DATABASE: NAME=PROW; NOTE=PROW 2:85-89(2001);
CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1644341535_g.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; J05550; AAA59868.1; -
DR EMBL; M93221; AAA60389.1; -
DR EMBL; M93192; AAA60389.1; JOINED.
DR EMBL; M93193; AAA60389.1; JOINED.
DR EMBL; M93194; AAA60389.1; JOINED.
DR EMBL; M93195; AAA60389.1; JOINED.
DR EMBL; M93196; AAA60389.1; JOINED.
DR EMBL; M93197; AAA60389.1; JOINED.
DR EMBL; M93198; AAA60389.1; JOINED.
DR EMBL; M93199; AAA60389.1; JOINED.
DR EMBL; M93200; AAA60389.1; JOINED.
DR EMBL; M93201; AAA60389.1; JOINED.
DR EMBL; M93202; AAA60389.1; JOINED.
DR EMBL; M93203; AAA60389.1; JOINED.
DR EMBL; M93204; AAA60389.1; JOINED.
DR EMBL; M93205; AAA60389.1; JOINED.
DR EMBL; M93206; AAA60389.1; JOINED.
DR EMBL; M93207; AAA60389.1; JOINED.
DR EMBL; M93208; AAA60389.1; JOINED.
DR EMBL; M93209; AAA60389.1; JOINED.
DR EMBL; M93210; AAA60389.1; JOINED.
DR EMBL; M93211; AAA60389.1; JOINED.
DR EMBL; M93212; AAA60389.1; JOINED.
DR EMBL; M93213; AAA60389.1; JOINED.
DR EMBL; M93214; AAA60389.1; JOINED.
DR EMBL; M93215; AAA60389.1; JOINED.
DR EMBL; M93216; AAA60389.1; JOINED.
DR EMBL; M93217; AAA60389.1; JOINED.
DR EMBL; M93218; AAA60389.1; JOINED.
DR EMBL; M93219; AAA60389.1; JOINED.
DR EMBL; M93220; AAA60389.1; JOINED.
DR PIR; A36563; A36563.
DR PDB; 1EGG; 30-AUG-00.
DR PDB; 1EGI; 30-AUG-00.
DR Genew; HGNC:7228; MRC1.
DR MIM; 153618; -
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0005537; F:mannose binding activity; TAS.
DR GO; GO:0004872; F:receptor activity; TAS.
DR GO; GO:0006898; P:receptor mediated endocytosis; TAS.
DR InterPro; IPR002353; AntifreezeII.
DR InterPro; IPR000562; FN_TypeII.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin_c; 8.
DR Pfam; PF00652; Ricin_B_lectin; 2.
DR PRINTS; PR00013; FNYPEII.
DR PRINTS; PR00356; ANTIFREEZEII.
DR ProDom; PD000995; FN_TypeII; 1.
DR SMART; SM00034; CLECT; 8.

```

DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; 6.
DR PROSITE; PS00041; C-TYPE LECTIN 2; 8.
DR PROSITE; PS00023; FIBRONECTIN 2; 1.
DR PROSITE; PS00231; RICIN B LECTIN; 1.
KW Receptor; Signal; Calcium-binding; Transmembrane; Repeat;
KW Glycoprotein; Endocytosis; Macrophage; Lectin; Antigen; 3D-structure.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 1456 MACROPHAGE MANNOSE RECEPTOR.
FT DOMAIN 19 1383 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1384 1411 POTENTIAL.
FT DOMAIN 1412 1456 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 22 142 RICIN B-TYPE LECTIN.
FT DOMAIN 157 212 FIBRONECTIN TYPE-II.
FT DOMAIN 216 344 C-TYPE LECTIN 1 (LONG FORM).
FT DOMAIN 360 490 C-TYPE LECTIN 2 (LONG FORM).
FT DOMAIN 502 629 C-TYPE LECTIN 3 (LONG FORM).
FT DOMAIN 644 781 C-TYPE LECTIN 4 (LONG FORM).
FT DOMAIN 805 926 C-TYPE LECTIN 5 (LONG FORM).
FT DOMAIN 943 1083 C-TYPE LECTIN 6 (LONG FORM).
FT DOMAIN 1100 1216 C-TYPE LECTIN 7 (LONG FORM).
FT DOMAIN 1228 1359 C-TYPE LECTIN 8 (LONG FORM).
FT DISULFID 646 659
FT DISULFID 680 777
FT DISULFID 753 769
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 930 930 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1160 1160 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1205 1205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1311 1311 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 648 649
FT STRAND 651 652
FT TURN 654 655
FT STRAND 658 663
FT HELIX 667 669
FT STRAND 671 671
FT HELIX 673 684
FT TURN 686 687
FT STRAND 693 705
FT TURN 706 707
FT STRAND 709 710
FT STRAND 712 718
FT TURN 723 724
FT STRAND 727 727
FT TURN 729 730
FT STRAND 733 733
FT TURN 741 742
FT HELIX 746 748
FT STRAND 752 757
FT TURN 758 761
FT STRAND 764 768
FT TURN 769 770
FT STRAND 773 780
FT TURN 781 782
SQ SEQUENCE 1456 AA; 166011 MW; 2645EAF3C576AE53 CRC64;

Query Match 16.5%; Score 182; DB 1; Length 1456;
Best Local Similarity 25.5%; Pred. No. 3.4e-08;
Matches 52; Conservative 37; Mismatches 61; Indels 54; Gaps 8;
QY 18 YKVIVFHDTSRLNFEAEACRRDGGQVLVSSEDEOKLIEKFIENLLPSDGDWIGLR 77
DB 807 YKDYQYYSKESKETWDMNARAFCKRFGDLVLSQSESEKFLWKYV-NRNDQSAFYIGLL 865
QY 78 FREEQKSNSTACQDLYAWTDGSIQFRNRYWDEPSCGS--EVCVVMYHQPAPAGIGGPY 135
DB 866 ISLDKK-----FAMWDGSKVDYVSWATGEPNANEDENCVTMY-----SNSGF----- 908

QY 136 MFQWDDRCNMKNFNICKYDEK----PAVPSREAGBETELTTPVLPEBQF----- 184
DB 909 ---WWDINGCYPNAFICQHRHSSINATTWMP-----TMPSPVSCCKEGMFPYSN 954
QY 185 -----EDAKYTKESREAAAL 199
DB 955 KCFKIFGFMEERKNWQAEARKACI 978
RESULT 4
PGCN_MOUSE
ID PGCN_MOUSE STANDARD; PRT; 1268 AA.
AC P55066;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
OS CSPG3 OR NCAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/c; TISSUE=Brain;
RX MEDLINE=96039250; PubMed=7490074;
RA Rauch U., Gripe B., Kulbe G., Arnold-Ammer I., Beier D.,
RT Paesler R.;
RL "Structure and chromosomal localization of the mouse neurocan gene.";
RL Genomics 28:405-410(1995).
CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC development by binding to neural cell adhesion molecules (NG-CAM
CC and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC acid.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X84727; CAA59216.1; -
DR PIR; S52781; S52781.
DR HSP; P00740; IEDM.
DR MGD; MGI:104694; Cspg3.
DR InterPro; IPR002353; AntifreezeII.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR PRINTS; PR00356; ANTIFREEZEII.
DR ProDom; PD000918; Link; 2.

```

DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG_LINK; 1.
DR PROSITE; PS01241; LINK; 2.
DR Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lentin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1268
FT DOMAIN 37 157
FT DOMAIN 158 253
FT DOMAIN 259 355
FT DOMAIN 360 996
FT DOMAIN 998 1034
FT DOMAIN 1036 1165
FT DOMAIN 1166 1224
FT DISULFID 58 139
FT DISULFID 181 252
FT DISULFID 205 226
FT DISULFID 279 354
FT DISULFID 303 324
FT DISULFID 364 975
FT DISULFID 969 984
FT DISULFID 986 995
FT DISULFID 1040 1051
FT DISULFID 1068 1160
FT DISULFID 1136 1152
FT DISULFID 1167 1210
FT DISULFID 1196 1223
FT CARBOHYD 121 121
FT CARBOHYD 339 339
FT CARBOHYD 742 742
FT CARBOHYD 978 978
FT CARBOHYD 1175 1175
SQ SEQUENCE 1268 AA; 137200 MW; 30148E8E20A2FAEC CRC64;

Query Match 16.1%; Score 177; DB 1; Length 1268;
Best Local Similarity 31.5%; Pred. No. 7.9e-08;
Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;

QY 14 QPCYKVIYFHTSRRLNFEAKACRRDGGQLVSISEDEQKLIKFIENLLPSDGDWF 73
Db 1048 QGHCYR--YF---AHRRAWEDAERDCRRAGHLTSVHSPEEHKFINSF-----GHENSW 1096
QY 74 IGLRRREKQSNSTACQDLVATWDGSIQFNNWVDEPS---CGSEVCVVMYHQPSAPAG 130
Db 1097 IGLNDRIVERD-----FQWTDNTGLQVENREKQPDNFFAGGEDCVVMVAHESG--- 1145
QY 131 IGGPYMFQWMDRNCNKNPFCK 153
Db 1146 -----RWNDVPCYNLPVCK 1161

RESULT 5
PGCN_HUMAN
ID PGCN_HUMAN STANDARD; PRT; 1321 AA.
AC 014594; Q9UPK6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
GN CSPG3 OR NCAN OR NEUR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99013874; PubMed=9795216;
RA Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Lennon G.G.;
RT "Characterization of the human neurocan gene, CSPG3.";
RL Gene 221:199-205(1998).
RN [2]
RP SEQUENCE OF 1-990 AND 1007-1321 FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Garnea J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Tranheim M., Amico-Keller G.,
RA Cosfield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krobayashi A., Olsen A.S., Carrano A.V.;
RA Krobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of an ~1 Mb region containing the MEF2B gene in
    19p12.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC development by binding to neural cell adhesion molecules (NC-CAM
CC and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC acid.
CC -!- TISSUE SPECIFICITY: Brain.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF026347; AAC80576.1; -
CC EMBL; AC003110; AAB86655.1; -
CC EMBL; AC005254; AAC25581.1; -
CC HSP; P00740; IEDM
CC MIM; 600826; -
CC DR InterPro; IPR000152; Asx_hydroxyl.
CC DR InterPro; IPR000742; EGF_2.
CC DR InterPro; IPR001881; EGF_Ca.
CC DR InterPro; IPR006209; EGF_like.
CC DR InterPro; IPR007110; IG-like.
CC DR InterPro; IPR003599; IG.
CC DR InterPro; IPR003006; Ig_MHC.
CC DR InterPro; IPR001304; Lentin_C.
CC DR InterPro; IPR000538; Link.
CC DR InterPro; IPR000436; Sushi_SCR_CCP.
CC DR Pfam; PF00008; EGF_2.
CC DR Pfam; PF00047; Ig; 1.
CC DR Pfam; PF00059; lectin_c; 1.
CC DR Pfam; PF00084; sushi; 1.
CC DR Pfam; PF00193; Xlink; 2.
CC DR PRINTS; PR01265; LINKMODULE.
CC DR ProDom; PD000918; Link; 2.
CC DR SMART; SM00032; CCP; 1.
CC DR SMART; SM00034; CLECT; 1.
CC DR SMART; SM00179; EGF_CA; 1.
CC DR SMART; SM00409; IG; 1.
CC DR SMART; SM00445; LINK; 2.
CC DR PROSITE; PS00010; ASX_HYDROXYL; 1.
CC DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
CC DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
CC DR PROSITE; PS00022; EGF_1; 3.
CC DR PROSITE; PS00022; EGF_2; 1.
CC DR PROSITE; PS01186; EGF_2; 1.

```

```
DR PROSITE; PS01187; EGF CA; 1.
DR PROSITE; PS0835; IG-LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT CHAIN 1 22 POTENTIAL.
FT CHAIN 23 1321 NEUCAN CORE PROTEIN.
FT DOMAIN 38 153 IG-LIKE V-TYPE.
FT DOMAIN 159 254 LINK 1.
FT DOMAIN 260 356 LINK 2.
FT DOMAIN 1008 1044 EGF-LIKE 1.
FT DOMAIN 1046 1082 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1084 1213 C-TYPE LECTIN.
FT DOMAIN 1214 1272 SUSHI.
FT DISULFID 59 140 BY SIMILARITY.
FT DISULFID 182 253 BY SIMILARITY.
FT DISULFID 206 227 BY SIMILARITY.
FT DISULFID 280 355 BY SIMILARITY.
FT DISULFID 304 325 BY SIMILARITY.
FT DISULFID 1012 1023 BY SIMILARITY.
FT DISULFID 1017 1032 BY SIMILARITY.
FT DISULFID 1034 1043 BY SIMILARITY.
FT DISULFID 1088 1099 BY SIMILARITY.
FT DISULFID 1116 1208 BY SIMILARITY.
FT DISULFID 1184 1200 BY SIMILARITY.
FT DISULFID 1215 1258 BY SIMILARITY.
FT DISULFID 1244 1271 BY SIMILARITY.
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1223 1223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1234 1234 N -> Y (IN REF. 2).
FT CONFLICT 1254 1254 V -> A (IN REF. 2).
FT CONFLICT 1282 1282 G -> R (IN REF. 2).
SQ SEQUENCE 1321 AA; 142972 MW; 2EF47F823DB98088 CRC64;

Query Match 16.0%; Score 176; DB 1; Length 1321;
Best Local Similarity 31.5%; Pred. No. 1e-07;
Matches 45; Conservative 18; Mismatches 48; Indels 32; Gaps 6;

QY 14 QPCYKVIYFHTSRNLNFEAEKACRRGGQGLVSTSEDEQKLEKFTENLLPSDGF 73
DB 1096 QGHYR--YF--AHRRADAEKDCRRSRGHLTSVHSPEHSFNSF-----GHENTW 1144
QY 74 IGLRRREKQSNSTACQDLVATDGSISOFRNWYVDPS---CGSEVCVMYHQSPAPAG 130
DB 1145 IGLNDRIVERD-----FQMTDNTGLOFENWRENQPDNFFAGGEDCVMVAHESG----1193
QY 131 IGGPYMFQWDRRCNMKNPFICK 153
DB 1194 -----RWNDVPCNVNLPYVCK 1209

RESULT 6
PGCV_BOVIN
ID PGCV_BOVIN STANDARD; PRT; 3381 AA.
AC P81282; O77609; O77610; O77611; O77612;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP).
GN CP52.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS VO; V1; V2 AND V3).
RC TISSUE=Forebrain;
RX MEDLINE=98288320; PubMed=9624174;
```

```
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
RL J. Biol. Chem. 273:15758-15764 (1998).
RN [2]
RP SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331
RP AND 342-348.
RC TISSUE=Spinal cord;
RX MEDLINE=92062692; PubMed=1720020;
RA Perides G., Biviano F., Bignami A.;
RT "Interaction of a brain extracellular matrix protein with hyaluronic
RT acid.";
RL Biochim. Biophys. Acta 1075:248-258 (1991).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBUNIT: Interacts with FBLN1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=P81282-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=P81282-2; Sequence=VSP_003078, VSP_003079;
CC Name=V2;
CC IsoId=P81282-3; Sequence=VSP_003080;
CC Name=V3;
CC IsoId=P81282-4; Sequence=VSP_003081, VSP_003081;
CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed
CC in the central nervous system, and in a number of mesenchymal and
CC epithelial tissues; the major isoform V2 is restricted to the
CC central nervous system.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (By similarity).
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN/PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-aib.ch/announce/
CC or send an email to license@isb-aib.ch).
CC -----
DR EMBL; AF060456; AAC24358.1; -
DR EMBL; AF060457; AAC24359.1; -
DR EMBL; AF060458; AAC24360.1; -
DR EMBL; AF060459; AAC24361.1; -
DR PIR; T14274; T14274.
DR PIR; T42389; T42389.
DR HSSP; P01132; LEPO.
DR InterPro; IPR00152; Asx_hydroxyl.
DR InterPro; IPR00742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR00538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; Lectin_C; 1.
DR Pfam; PF00084; sushi; 1.
```

| | | | | |
|---|--|---|------------|---|
| FT | VARSPLIC | 350 | 1336 | /FTid=VSP_003078. Missing (in isoform V1). |
| FT | VARSPLIC | 350 | 1336 | /FTid=VSP_003079. |
| FT | VARSPLIC | 1337 | 3074 | Missing (in isoform V2). |
| FT | VARSPLIC | 350 | 3074 | /FTid=VSP_003080. |
| FT | VARSPLIC | 350 | 3074 | Missing (in isoform V3). |
| FT | VARSPLIC | 350 | 3074 | /FTid=VSP_003081. |
| FT | CONFLICT | 25 | 25 | MISSING (IN REF. 2). |
| FT | CONFLICT | 51 | 51 | MISSING (IN REF. 2). |
| FT | CONFLICT | 89 | 89 | N -> D (IN REF. 2). |
| FT | CONFLICT | 96 | 96 | O -> D (IN REF. 2). |
| FT | CONFLICT | 346 | 346 | C -> R (IN REF. 2). |
| SQ | SEQUENCE | 3381 AA, | 369984 MW; | F09716FA778D459 CRC64; |
| | | | | |
| Query Match 15.9%; Score 174.5; DB 1; Length 3381; | | | | |
| Best Local Similarity 25.5%; Pred. No. 4.4e-07; | | | | |
| Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11; | | | | |
| | | | | |
| QY | 14 | QRPCYKVIYFHDTSRRRLNFEAEACRRDGGQLVSIESDEQKLIEKFLENLPSDGF-- | 72 | |
| Db | 3162 | GQCQCYK--YF---AHRRTWDAARECRLOGAHLTSLSHBEQMFAVRV-----GHDVQ | 3209 | |
| | | | | |
| QY | 73 | WIGLRREEQKSSTACODLIYATDGSGISOFRNWYDEP-----SCSEVCVMYHQSAP | 128 | |
| Db | 3210 | WIGL-----NDXMFEDHFRWTGDSLTQYENRPNQDPSPFSTGEDCVIWHENG--- | 3259 | |
| | | | | |
| QY | 129 | AGIGGPYMFQWDDRCNMKNFFICKYS----DEKPAPVPREAGE----- | 169 | |
| Db | 3260 | -----QWNVPCNYHLTYTCCKGTACGGPPVVENAKTFGMKPRYEINSLRYHC | 3310 | |
| | | | | |
| QY | 170 | -----EPELTT-----PVL-----PEETQEEDAKKTPKESRAALN | 200 | |
| Db | 3311 | KDGIQIRHLPTIRCLGNRWAMPKITCLNPASAYQRTYSKYFKNSSSAXDN | 3361 | |
| | | | | |
| RESULT 7 | | | | |
| PGCN RAT | ID | PGCN RAT | STANDARD; | PRT; 1257 AA. |
| AC | P5067; | | | |
| DT | 01-OCT-1996 | (Rel. 34, Created) | | |
| DT | 01-OCT-1996 | (Rel. 34, Last sequence update) | | |
| DT | 15-SEP-2003 | (Rel. 42, Last annotation update) | | |
| DE | Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3) | | | |
| DE | (245 kDa early postnatal core glycoprotein) [Contains: 150 kDa adult | | | |
| DE | core glycoprotein]. | | | |
| GN | CSPG3 OR NCAN. | | | |
| OS | Rattus norvegicus (Rat). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | | | |
| OX | NCBI_TaxID=10116; | | | |
| ON | [1] | | | |
| RP | SEQUENCE FROM N.A., AND PARTIAL SEQUENCE. | | | |
| RC | STRAIN=Sprague-Dawley; TISSUE=Brain; | | | |
| RX | MEDLINE=92406907; PubMed=1326557; | | | |
| RA | Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.; | | | |
| RA | "Cloning and primary structure of neurocan, a developmentally | | | |
| RT | regulated, aggregating chondroitin sulfate proteoglycan of brain."; | | | |
| RL | J. Biol. Chem. 267:19536-19547(1992). | | | |
| RN | [2] | | | |
| RP | CHARACTERIZATION. | | | |
| RX | MEDLINE=94230574; PubMed=7513709; | | | |
| RA | Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K., | | | |
| RA | Margolis R.U., Grumet M.; | | | |
| RT | "The neuronal chondroitin sulfate proteoglycan neurocan binds to the | | | |
| RT | neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits | | | |
| RT | neuronal adhesion and neurite outgrowth." | | | |
| RL | J. Cell Biol. 125:669-680(1994). | | | |
| CC | -1- FUNCTION: May modulate neuronal adhesion and neurite growth during | | | |
| CC | development by binding to neural cell adhesion molecules (NG-CAM | | | |
| CC | and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic | | | |
| CC | acid. | | | |
| CC | -1- TISSUE SPECIFICITY: EARLY POSTNATAL AND ADULT BRAIN; NOT EXPRESSED | | | |
| CC | IN KIDNEY, LUNG, LIVER AND MUSCLE. | | | |

[4]
 RN SEQUENCE OF 2535-2738 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Lung;
 RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;
 RT "Molecular cloning and characterization of two developmentally
 regulated genes in rat lung.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May play a role in intercellular signaling and in
 connecting cells with the extracellular matrix. May take part in
 the regulation of cell motility, growth and differentiation. Binds
 hyaluronic acid.
 CC -1- SUBUNIT: Interacts with FBLN1 (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Additional isoforms seem to exist;
 CC Names=V0;
 CC IsoId=Q9ERB4-1; Sequence=Displayed;
 CC Names=V3;
 CC IsoId=Q9ERB4-2; Sequence=VSP_003091;
 CC Names=Vint;
 CC IsoId=Q9ERB4-3; Sequence=VSP_003092;
 CC TISSUE SPECIFICITY: In kidney is expressed in the papillary area,
 CC but not in glomeruli.
 CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development
 CC (By similarity).
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -1- SIMILARITY: Contains 2 link domains.
 CC -1- SIMILARITY: Contains 2 EGF-like domains.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. Usage by and for commercial
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AF062402; AAC40166.1; -;
 CC EMBL; U75306; AAB51125.1; -;
 CC EMBL; AF084544; AAD48544.1; -;
 CC EMBL; AF072892; AAC26116.1; -;
 CC EMBL; AY007691; AAG16631.1; -;
 CC HSSP; P01132; 1EPG.
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR003599; Ig.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR000538; Link.
 CC InterPro; IPR000436; Sushi_SCR_OCC.
 CC PRINTS; PR01265; LINKMODULE.
 CC PRINTS; PR00356; ANTIFREEZEII.
 CC ProDom; PD000918; Link; 2.
 CC SMART; SM00032; CCP; 1.
 CC SMART; SM00034; CLECT; 1.
 CC SMART; SM00179; EGF_CA; 1.
 CC SMART; SM00409; IG; 1.
 CC SMART; SM00445; LINK; 2.
 CC PROSITE; PS00010; ASX_HYDROXYL; 1.
 CC PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
 CC PROSITE; PS00441; C_TYPE_LLECTIN_2; 1.
 CC PROSITE; PS00022; EGF_1; 2.
 CC PROSITE; PS01186; EGF_2; 1.
 CC PROSITE; PS01187; EGF_CA; 1.
 CC PROSITE; PS00835; IG_LIKE; 1.
 CC PROSITE; PS01241; LINK; 2.
 CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;

KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2738 VERSICAN CORE PROTEIN.
 FT NON_CONS 348 349
 FT DOMAIN 21 146 IG-LIKE V-TYPE.
 FT DOMAIN 167 244 LINK 1.
 FT DOMAIN 265 346 LINK 2.
 FT DOMAIN <349 695 GAG-ALPHA
 (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
 FT DOMAIN 696 2431 GAG-BETA.
 FT DOMAIN 2431 2467 EGF-LIKE 1.
 FT DOMAIN 2469 2505 EGF-LIKE 2, CALCIUM-BINDING.
 FT DOMAIN 2518 2632 C-TYPE LECTIN.
 FT DOMAIN 2637 2695 SUSHI.
 FT DISULFID 44 130 BY SIMILARITY.
 FT DISULFID 172 243 BY SIMILARITY.
 FT DISULFID 196 217 BY SIMILARITY.
 FT DISULFID 270 345 BY SIMILARITY.
 FT DISULFID 294 315 BY SIMILARITY.
 FT DISULFID 2435 2446 BY SIMILARITY.
 FT DISULFID 2440 2455 BY SIMILARITY.
 FT DISULFID 2473 2484 BY SIMILARITY.
 FT DISULFID 2478 2493 BY SIMILARITY.
 FT DISULFID 2495 2504 BY SIMILARITY.
 FT DISULFID 2511 2522 BY SIMILARITY.
 FT DISULFID 2539 2631 BY SIMILARITY.
 FT DISULFID 2607 2623 BY SIMILARITY.
 FT DISULFID 2638 2681 BY SIMILARITY.
 FT DISULFID 2667 2694 BY SIMILARITY.
 FT CARBOHYD 57 57 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 692 692 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 758 758 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 805 805 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1257 1257 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1435 1435 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1633 1633 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1660 1660 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1684 1684 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1738 1738 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 1848 1848 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 2004 2004 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 2409 2409 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 2711 2711 N-LINKED (GLCNAC.) (POTENTIAL).
 FT CARBOHYD 2721 2721 N-LINKED (GLCNAC.) (POTENTIAL).
 FT VARSPLIC 349 2431 Missing (in isoform V3).
 FT VARSPLIC 2697 2738 /FTId=VSP_003091.
 FT VARSPLIC 2697 2738 PSAYORTYTKYKLLKSSSVKNSINTSKHEHRWSRWQETR
 FT R -> RKWSFRKNGQPCFNKY (in isoform Vint).
 FT /FTId=VSP_003092.
 FT CONFLICT 2535 2539 AEREC -> NSARG (IN REF. 4).
 FT SEQUENCE 2738 AA; 300004 MW; 12CA626D58BD8C6A CRC64;
 SQ

Query Match 15.8%; Score 174; DB 1; Length 2738;
 Best Local Similarity 28.5%; Pred. No. 3.8e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
 QY 14 QRPCKYVYFHDTSRLNFEAEKARRDGGOLVSTESDEQKLEKFTIENLLPSGDGF- 72
 DB 2519 QQQCYK--YF-----AHRRTWDAARECRQLQGAHLTSLSHBQMFNVRV-----GHDYQ 2566
 QY 73 WIGLRREEKQSNSTACQDLVATWDGSIQFANWTVDEP-----SCQSEVVCVVMYHPSAP 128
 DB 2567 MIGL-----NDKMFHDFRWTDGSAALQYENWPNQDPSPFSAGEDCVIWHENG-- 2616
 QY 129 AGIGPYMFQNDNDRCNMKNPFICKYS-----DEKPAVPSREAGE 169
 DB 2617 -----QWMDVPCNYHLTYTCKKGTVACQPPVVENAKTFGK 2652

RESULT 9
 PGCV MOUSE
 ID PGCV MOUSE STANDARD; PRT; 3358 AA.
 AC Q62059; Q62058; Q9CU00;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Versican core protein precursor (Large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
 GN CSPG2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS V0; V1 AND V2).
 RC STRAIN=C57BL/6; and Swiss Webster; TISSUE=Brain;
 RX MEDLINE=95181355; PubMed=7822336;
 RA Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;
 RT "Multiple forms of mouse PG-M, a large chondroitin sulfate
 RT proteoglycan generated by alternative splicing.";
 RL J. Biol. Chem. 270:958-965(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM V3).
 RC STRAIN=C57BL/6;
 RX MEDLINE=95181355; PubMed=7876137;
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
 RT without a chondroitin sulfate attachment in region in mouse and human
 RT tissues.";
 RL J. Biol. Chem. 270:3914-3918(1995).
 RN [3]
 RP SEQUENCE OF 1-1692 FROM N.A. (ISOFORM V1).
 RC STRAIN=C57BL/6J; TISSUE=Skin;
 RX MEDLINE=21085660; PubMed=1117851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gassner T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gusninch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [4]
 RP INTERACTION WITH FBLN1.
 RX PubMed=10400671;
 RA Asperger A., Adam S., Kostka G., Timpl R., Heinemann D.;
 RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and
 RT versican.";
 RL J. Biol. Chem. 274:20444-20449(1999).
 CC -!- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -!- SUBUNIT: Interacts with FBLN1.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Comment=Additional isoforms seem to exist;
 CC Name=V0;
 CC IsoId=Q62059-1; Sequence=Displayed;

CC CC Name=V1;
 CC IsoId=Q62059-2; Sequence=VSP_003087, VSP_003088;
 CC Name=V2;
 CC IsoId=Q62059-3; Sequence=VSP_003089;
 CC Name=V3;
 CC IsoId=Q62059-4; Sequence=VSP_003087, VSP_003090;
 CC -!- TISSUE SPECIFICITY: V2 is found only in brain.
 CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D1263; BAA03796.1; -;
 CC EMBL; D28599; -; NOT ANNOTATED_CDS.
 CC EMBL; D32040; BAA06802.1; -;
 CC EMBL; AK014525; BAB29411.1; -;
 CC HSP; F01132; IEPG.
 CC MGD; MGI:102889; Cpg92.
 CC InterPro; IPR000152; Abx hydroxyl.
 CC InterPro; IPR00742; EGF_2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF_Like.
 CC InterPro; IPR007110; Ig-Like.
 CC InterPro; IPR003599; Ig.
 CC InterPro; IPR003006; Ig MHC.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR000538; Link.
 CC InterPro; IPR000436; Sushi_SCR_CCP.
 CC Pfam; PF00008; EGF; 2.
 CC Pfam; PF00047; Ig; 1.
 CC Pfam; PF00059; lectin_c; 1.
 CC Pfam; PF00084; sushi; 1.
 CC Pfam; PF00193; Xlink; 2.
 CC PRINTS; PR01265; LINKMODULE.
 CC ProDom; PD000918; Link; 2.
 CC SMART; SM00032; CCP; 1.
 CC SMART; SM00034; CLECT; 1.
 CC SMART; SM00179; EGF_CA; 1.
 CC SMART; SM00409; IG_1.
 CC SMART; SM00445; LINK; 2.
 CC PROSITE; PS00010; ASX_HYDROXYL; 1.
 CC PROSITE; PS00615; C-TYPE LECTIN 1; 1.
 CC PROSITE; PS00041; C-TYPE LECTIN 2; 1.
 CC PROSITE; PS00022; EGF 1; 2.
 CC PROSITE; PS01186; EGF 2; 1.
 CC PROSITE; PS01187; EGF_CA; 1.
 CC PROSITE; PS00835; IG_LIKE; 1.
 CC PROSITE; PS01241; LINK; 2.
 CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 CC Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 CC Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 3358 VERSICAN CORE PROTEIN.
 FT DOMAIN 21 146 IG-LIKE V-TYPE.
 FT DOMAIN 167 244 LINK 1.
 FT DOMAIN 265 346 LINK 2.
 FT DOMAIN 348 1308 GAG-ALPHA
 FT (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
 FT DOMAIN 1309 3052 GAG-BETA.
 FT DOMAIN 3052 3088 EGF-LIKE 1.
 FT DOMAIN 3090 3126 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 3139 3253 C-TYPE LECTIN.

```

FT DOMAIN 3258 3316 SUSHI.
FT DISULFID 44 130 BY SIMILARITY.
FT DISULFID 172 243 BY SIMILARITY.
FT DISULFID 196 217 BY SIMILARITY.
FT DISULFID 270 333 BY SIMILARITY.
FT DISULFID 294 315 BY SIMILARITY.
FT DISULFID 3056 3067 BY SIMILARITY.
FT DISULFID 3061 3076 BY SIMILARITY.
FT DISULFID 3078 3087 BY SIMILARITY.
FT DISULFID 3094 3105 BY SIMILARITY.
FT DISULFID 3099 3114 BY SIMILARITY.
FT DISULFID 3116 3125 BY SIMILARITY.
FT DISULFID 3132 3143 BY SIMILARITY.
FT DISULFID 3160 3252 BY SIMILARITY.
FT DISULFID 3228 3244 BY SIMILARITY.
FT DISULFID 3259 3302 BY SIMILARITY.
FT DISULFID 3288 3315 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 914 914 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1305 1305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1372 1372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1679 1679 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2054 2054 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2244 2244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2362 2362 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2627 2627 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3030 3030 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3332 3332 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3342 3342 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 348 348 P -> R (in isoform V1 and isoform V3).
FT VARSPLIC 349 1308 /FTid=VSP 003087.
FT VARSPLIC 1309 3052 Missing (in isoform V1).
FT VARSPLIC 349 3052 /FTid=VSP 003088.
FT VARSPLIC 349 3052 Missing (in isoform V2).
FT VARSPLIC 349 3052 Missing (in isoform V3).
FT CONFLICT 126 126 A -> G (IN REF. 3).
FT CONFLICT 348 348 MISSING (IN REF. 3).
FT CONFLICT 1658 1658 I -> T (IN REF. 3).
FT CONFLICT 1674 1680 TWNSNS -> QFGIQTA (IN REF. 3).
SQ SEQUENCE 3358 AA; 366938 MW; 071B800268C0762D CRC64;

Query Match
Best Local Similarity 15.8%; Score 174; DB 1; Length 3358;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

QY 14 QRCYKVIYFHTSRLNFEAEKACRDGGQLVSIIESEDEQKLEKFIENLLPSDGF- 72
Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
3140 QGQCYK--YF---AHRRTWDAERECRLQGAHLTSLTSHHEQMFINRV-----CHDQ 3187

QY 73 WGLRREKQSNSTACQDIYATDGSISQFRNYYVDEP-----SCSGVCVVMYHOPSAP 128
Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
3188 WIGL-----NDKMFEDHFRDTGSLQYENWRNPQDSFFSAGEDCWIWHENG-- 3237

QY 129 AGIGGPFQWDDRCNKNKFKYS----DEKPAVPSREAGE 169
Db ||||| : : : : : ||||| : : : : : ||||| : : : : :
3238 -----QWNDVPCNYHLTCTCKGTACQPPVVENAKTFGK 3273

RESULT 10
PGCV HUMAN
AC P13611; P20754; Q13010; Q13189; Q15123; Q9UNW5;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)

```

```

DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP).
GN CSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM V0).
RX MEDLINE=95105188; PubMed=7528742;
RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
RT "Characterization of the complete genomic structure of the human
RT versican gene and functional analysis of its promoter.";
RL J. Biol. Chem. 269:32999-33008(1994).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM V1).
RX TISSUE=Placenta;
RA Zimmermann D.R., Ruoslahti E.;
RT "Multiple domains of the large fibroblast proteoglycan, versican.";
RL EMBO J. 8:2975-2981(1989).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM V2).
RX TISSUE=Glial tumor;
RA Dours-Zimmermann M.T., Zimmermann D.R.;
RT "A novel glycosaminoglycan attachment domain identified in two
RT alternative splice variants of human versican.";
RL J. Biol. Chem. 269:32992-32998(1994).
RN [4]
RP SEQUENCE OF 2711-3396 FROM N.A.
RX TISSUE=Lung fibroblast;
RA MEDLINE=88007514; PubMed=2820964;
RT Krusius T., Gehlsen K.R., Ruoslahti E.;
RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
RT lectin-like and growth factor-like sequences.";
RL J. Biol. Chem. 262:13120-13125(1987).
RN [5]
RP SEQUENCE OF 251-347 FROM N.A.
RX MEDLINE=93122792; PubMed=1478664;
RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
RA McPherson J.D.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
RL J. Biol. Chem. 270:3914-3918(1995).
RN [7]
RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RX TISSUE=Aortic smooth muscle;
RA MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wight T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [8]
RP PARTIAL SEQUENCE.
RX TISSUE=Brain;
RA MEDLINE=89174663; PubMed=2466833;
RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
RT "Isolation and partial characterization of a glial
RT hyaluronate-binding protein.";
RL J. Biol. Chem. 264:5981-5987(1989).
RN [9]
RP TISSUE SPECIFICITY OF ISOFORMS.
RX MEDLINE=96213482; PubMed=8627343;

```

RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "Differential expression of versican isoforms in brain tumors.";
 RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
 CC -!- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -!- SUBUNIT: Interacts with FBLN1 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=5;
 CC Comment=Additional isoforms seem to exist;
 CC Name=V0;
 CC IsoId=P13611-1; Sequence=Displayed;
 CC Name=V1;
 CC IsoId=P13611-2; Sequence=VSP_003082, VSP_003083;
 CC Name=V2;
 CC IsoId=P13611-3; Sequence=VSP_003084;
 CC Name=V3;
 CC IsoId=P13611-4; Sequence=VSP_003082, VSP_003085;
 CC Name=Vint;
 CC IsoId=P13611-5; Sequence=VSP_003086;
 CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
 CC in normal brain, gliomas, medulloblastomas, schwannomas,
 CC neurofibromas, and meningiomas; V2 is restricted to normal brain
 CC and gliomas; V3 is found in all these tissues except
 CC medulloblastomas.
 CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U16306; AAA65018.1; -;
 DR EMBL; X15998; AAA34128.1; -;
 DR EMBL; S52488; AAB24878.1; -;
 DR EMBL; U26555; AAA67565.1; -;
 DR EMBL; D32039; BAA06801.1; -;
 DR EMBL; J02814; AAA36437.1; -;
 DR EMBL; AF084545; AAD48545.1; -;
 DR PIR; S06014; A60979.
 DR HSSP; P01132; 1EGF.
 DR Genew; HGNC:2464; CSPG2.
 DR MIM; 118661; -;
 DR GO; GO:0005578; C:extracellular matrix; TAS.
 DR GO; GO:0005204; F:chondroitin sulfate proteoglycan; TAS.
 DR GO; GO:0005540; F:hyaluronic acid binding activity; TAS.
 DR GO; GO:0008037; P:cell recognition; TAS.
 DR GO; GO:0007275; P:development; TAS.
 DR InterPro; IPRO00152; ASX hydroxyl.
 DR InterPro; IPRO00742; EGF_2.
 DR InterPro; IPRO01881; EGF_Ca.
 DR InterPro; IPRO06209; EGF-like.
 DR InterPro; IPRO07110; IG-like.
 DR InterPro; IPRO03599; IG.
 DR InterPro; IPRO03006; IG_MHC.
 DR InterPro; IPRO01304; Lectin_C.
 DR InterPro; IPRO00538; Link.
 DR InterPro; IPRO00436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00059; Lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.

DR ProDom; P0000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX HYDROXYL; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
 DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS050835; IG LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 20
 FT CHAIN 21 3396
 FT DOMAIN 21 146
 FT DOMAIN 167 244
 FT DOMAIN 265 346
 FT DOMAIN 348 1335
 FT DOMAIN 1336 3089
 FT DOMAIN 3089 3125
 FT DOMAIN 3127 3163
 FT DOMAIN 3176 3290
 FT DOMAIN 3295 3353
 FT DISULFID 44 130
 FT DISULFID 172 243
 FT DISULFID 196 217
 FT DISULFID 270 345
 FT DISULFID 294 315
 FT DISULFID 3093 3104
 FT DISULFID 3098 3113
 FT DISULFID 3115 3124
 FT DISULFID 3131 3142
 FT DISULFID 3136 3151
 FT DISULFID 3153 3162
 FT DISULFID 3169 3180
 FT DISULFID 3197 3289
 FT DISULFID 3265 3281
 FT DISULFID 3296 3339
 FT DISULFID 3325 3352
 FT CARBOHYD 57 57
 FT CARBOHYD 330 330
 FT CARBOHYD 615 615
 Query Match 15.8%; Score 174; DB 1; Length 3396;
 Best Local Similarity 28.5%; Pred No. 4.9e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
 QY 14 QRPCYKVIYFHDTSRRRLNFEAKACRRDGGQLVSISEDEQKLIETIENLLPSDGP- 72
 Db 3177 QGQCVK-YF---AHRRTWDAARECLQGAHLTSLSHSQMFVNRV-----GHDYQ 3224
 QY 73 WIGLRREKQSNSTACQDLYAWTDGSIQPRNRYTVDEP-----SCGSEVCVVMYHQPSAP 128
 Db 3225 WIGL-----NDKMFEDHFRWTDGSLQYENRPNQDSFFSAGEDCVVIWHENG-- 3274
 QY 129 AGIGGPFYMFOWNDRCNMKNFKYS-----DEKPAVPSRAEGE 169
 Db 3275 -----QNDVPCNVHLTYTCKGTVACGQPPVENAKTFGK 3310
 RESULT 11
 PGCV CHICK
 ID PGCV CHICK STANDARD; PRT; 3562 AA.
 AC Q90953; Q90945;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)

DR Versican core protein precursor (Large fibroblast proteoglycan)
 DR (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
 GN CS92.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS V0 AND V1).
 RC STRAIN=White leghorn; TISSUE=Limb bud;
 RX MEDLINE=93300846; PubMed=8314802;
 RA Shinomura T., Nishida Y., Ito K., Kimata K.;
 RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
 RT expressed during chondrogenesis in chick limb buds. Alternative
 RT spliced multiforms of PG-M and their relationships to versican.";
 RL J. Biol. Chem. 268:14461-14469(1993).
 CC -!- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=V0;
 CC IsoId=Q90953-1; Sequence=Displayed;
 CC Name=V1;
 CC IsoId=Q90953-2; Sequence=VSP_003093;
 CC -!- TISSUE SPECIFICITY: Prechondrogenic condensation area of
 CC developing limb buds.
 CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
 CC (by similarity).
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X60226; BAA42787.1; -;
 DR EMBL; D13542; BAA02742.1; -;
 DR PIR; A47171; A47171.
 DR HSP; P00740; LEDM.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR007110; Ig_Like.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00047; igf_1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.
 DR PRINTS; PR01265; LINKMODULE.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
 DR PROSITE; PS0041; C_TYPE_LLECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1_2;
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS00835; IG_LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 26
 FT CHAIN 27 3562
 FT DOMAIN 27 143
 FT DOMAIN 166 243
 FT DOMAIN 264 345
 FT DOMAIN 3254 3290
 FT DOMAIN 3292 3328
 FT DOMAIN 3341 3455
 FT DOMAIN 3460 3518
 FT DISULFID 44 129
 FT DISULFID 171 242
 FT DISULFID 195 216
 FT DISULFID 269 344
 FT DISULFID 293 314
 FT DISULFID 3258 3269
 FT DISULFID 3263 3278
 FT DISULFID 3280 3289
 FT DISULFID 3296 3307
 FT DISULFID 3301 3316
 FT DISULFID 3318 3327
 FT DISULFID 3334 3345
 FT DISULFID 3362 3454
 FT DISULFID 3430 3446
 FT DISULFID 3461 3504
 FT DISULFID 3490 3517
 FT CARBOHYD 163 163
 FT CARBOHYD 235 235
 FT CARBOHYD 329 329
 FT CARBOHYD 529 529
 FT CARBOHYD 709 709
 FT CARBOHYD 948 948
 FT CARBOHYD 1409 1409
 FT CARBOHYD 1479 1479
 FT CARBOHYD 1523 1523
 FT CARBOHYD 1530 1530
 FT CARBOHYD 1625 1625
 FT CARBOHYD 1751 1751
 FT CARBOHYD 1988 1988
 FT CARBOHYD 2088 2088
 FT CARBOHYD 2089 2089
 FT CARBOHYD 2507 2507
 FT CARBOHYD 2642 2642
 FT CARBOHYD 2679 2679
 FT CARBOHYD 2748 2748
 FT CARBOHYD 2762 2762
 FT CARBOHYD 3069 3069
 FT CARBOHYD 3194 3194
 FT CARBOHYD 3232 3232
 FT CARBOHYD 3545 3545
 FT VARSPLIC 485 1411
 FT Missing (in isoform V1).
 FT /FTid=VSP_003093.
 SQ SEQUENCE 3562 AA; 388078 MW; 9BC566E88C1602D2 CRC64;

Query Match 15.5%; Score 171; DB 1; Length 3562;
 Best Local Similarity 28.5%; Pred. No. 9.7e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

OY 14 QRPCKVTVYFHDTSRLNFEAKEACRRDGGOLVSTESDEQKLTKEPTIENLLPSGDF- 72
 Db 3342 QQCCK--YF---AHRTWDTAERECRLQGAHLTSLSHBEQVFNRI-----GHDIQ 3389
 OY 73 WIGLRREEKQSNSTACQDLVATWDGSIQSFRNVTYDEP-----SCGSEVCVVVHYQPSAP 128

```

Db 3390 WIGL-----NDKMFEDFRWTDGSPLOQYENRNQPDSPFSAGEDCVIIWHENG-- 3439
QY 129 AGIGGPMQWMDRCKMKNFICKYS-----DKPAVPSPREAGE 169
Db 3440 -----QWMDVPCNYHLTYTKKGTVACGGPPVVENAKTFGK 3475

RESULT 12
CD93_RAT
ID Q9ET61; Q9JIZ6; STANDARD; PRT; 643 AA.
AC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Complement component C1q receptor precursor (Complement component 1, q
DE subcomponent, receptor 1) (ClqR) (ClqR(p) (Clq/MBL/SPA receptor)
DE (CD93 antigen) (Cell surface antigen AA4).
GN ClqR1 OR CD93 OR ClqRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=PVG; TISSUE=Natural killer cells;
RX MEDLINE=20545218; PubMed=11093152;
RA Lovik G., Vaage J.T., Diesen E., Szpirer C., Ryan J.C., Rolstad B.;
RT "Characterization and molecular cloning of rat ClqR, a receptor on NK
RL Eur. J. Immunol. 30:3355-3362(2000).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Lung;
RX MEDLINE=20507883; PubMed=10934210;
RA Dean Y.D., McGreal E.P., Akatsu H., Gasque P.;
RT "Molecular and cellular properties of the rat AA4 antigen, a C-type
RT lectin-like receptor with structural homology to thrombomodulin."
RL J. Biol. Chem. 275:34382-34392(2000).
CC -!- FUNCTION: Receptor (or element of a larger receptor complex) for
CC Clq, mannose-binding lectin (MBL2) and pulmonary surfactant
CC protein A (SPA). May mediate the enhancement of phagocytosis in
CC monocytes and macrophages upon interaction with soluble defense
CC collagens. May play a role in intercellular adhesion.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Widely expressed. Highly expressed in lung and
CC heart. Expressed at lower level in brain, thymus, liver, spleen,
CC intestine, kidney, adrenal gland, muscle and testis. Expressed on
CC endothelial cells, platelets, undifferentiated monocytes and
CC circulating natural killer cells.
CC -!- PTM: N- and O-glycosylated (By similarity).
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 5 EGF-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF136537; AAG01572.1; -.
CC DR EMBL; AF160978; AAF80402.1; -.
CC HSP; P35555; 1EMN.
CC DR GO; GO:0016021; C:integral to membrane; ISS.
CC DR GO; GO:0004872; F:receptor activity; ISS.
CC DR GO; GO:0016337; P:cell-cell adhesion; ISS.
CC DR GO; GO:0042116; P:macrophage activation; ISS.
CC DR GO; GO:0006909; P:phagocytosis; ISS.
CC DR InterPro; IPR000152; AaX hydroxyl.
CC DR InterPro; IPR001881; EGF_Ca.
CC DR InterPro; IPR006209; EGF_Like.

```

```

DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00008; EGF_5.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
LR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX HYDROXYL; 3.
DR PROSITE; PS00615; C TYPE LECTIN 1; FALSE_NEG.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS01186; EGF_2_3.
DR PROSITE; PS01187; EGF_CA; 3.
KW Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
KW EGF-like domain; Lectin; Glycoprotein.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 643 COMPLEMENT COMPONENT C1Q RECEPTOR.
FT DOMAIN 24 571 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 572 592 POTENTIAL.
FT DOMAIN 593 643 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 31 173 C-TYPE LECTIN.
FT DOMAIN 257 298 EGF-LIKE 1.
FT DOMAIN 299 341 EGF-LIKE 2.
FT DOMAIN 342 381 EGF-LIKE 3.
FT DOMAIN 382 423 EGF-LIKE 4.
FT DOMAIN 424 462 EGF-LIKE 5.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 268 282 BY SIMILARITY.
FT DISULFID 284 297 BY SIMILARITY.
FT DISULFID 303 314 BY SIMILARITY.
FT DISULFID 308 325 BY SIMILARITY.
FT DISULFID 327 340 BY SIMILARITY.
FT DISULFID 346 355 BY SIMILARITY.
FT DISULFID 351 364 BY SIMILARITY.
FT DISULFID 366 380 BY SIMILARITY.
FT DISULFID 386 397 BY SIMILARITY.
FT DISULFID 393 406 BY SIMILARITY.
FT DISULFID 408 422 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 433 446 BY SIMILARITY.
FT DISULFID 448 461 BY SIMILARITY.
FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 417 417 E -> K (IN REF. 2).
SQ SEQUENCE 643 AA; 68781 MW; 9AE4C933AD943DB6 CRC64;

Query Match 14.9%; Score 164; DB 1; Length 643;
Best Local Similarity 25.5%; Pred. No. 5e-07;
Matches 49; Conservative 33; Mismatches 70; Indels 40; Gaps 9;

QY 8 VCGGTQRCYKVIYFHDTSRLNFEAEKACRRDGGQLVSISEDEQKLIKPIENLL- 66
Db 29 VCEG---TACYTAHW-----GKLSAAEAQHRCNENGGLATVKSEEARHVQEAQLLK 80
QY 67 ---PSD---GDWIGLRREKQSNSTACQDLYAWT-DGSIQFRNYY-VDEPSCGSEVC 118
Db 81 TKAPSETKIGKFMWIGLQREKGCCTYHDLPMKGFWSYWGGEEDTYSNWKASKSCISKRC 140
QY 119 VVMY-----HPSAPAGIGGPMYFQWMDRC-----NMKNFICKYSDSKPAVPSR 164
Db 141 VSLIDLSLKPSPHLP-----KWHESPCTPDAPGNSIEGLCFKPNFKOMCSPLA 191
QY 165 EAGGETELTTP 176
Db 192 LGPGQLTYTTP 203

RESULT 13
PGCA_BOVIN
ID PGCA_BOVIN STANDARD; PRT; 2364 AA.
AC P13608; P79117; Q28159;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSPCP).

```

GN AGCL.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Hering T.M., Kollar J., Huynh T.D.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 563-1056 FROM N.A.
RA MEDLINE=89380219; PubMed=2528543;
RX Antonsen P., Heinegaard D., Oldberg A.;
RN "The keratan sulfate-enriched region of bovine cartilage proteoglycan
RT consists of a consecutively repeated hexapeptide motif.";
RL J. Biol. Chem. 264:16170-16173(1989).
RN [3]
RP SEQUENCE OF 1609-2113 AND 2151-2364 FROM N.A.
RA MEDLINE=87270630; PubMed=3111460;
RX Oldberg A., Antonsen P., Heinegaard D.;
RN "The partial amino acid sequence of bovine cartilage proteoglycan,
RT deduced from a cDNA clone, contains numerous Ser-gly sequences
RL arranged in homologous repeats.";
RL Biochem. J. 243:255-259(1987).
RN [4]
RP SEQUENCE OF 2114-2150 FROM N.A.
RA TISSUE=Cartilage;
RX MEDLINE=93352525; PubMed=8349621;
RN Fuelop C., Walcz E., Valyon M., Glant T.T.;
RT "Expression of alternatively spliced epidermal growth factor-like
RL domains in aggrecans of different species. Evidence for a novel
RL module.";
RL J. Biol. Chem. 268:17377-17383(1993).
RN [5]
RP PARTIAL SEQUENCE.
RX MEDLINE=85027710; PubMed=6489519;
RN Perin J.-P., Bonnet F., Jolles J., Jolles P.;
RT "Sequence data concerning the protein core of the cartilage
RL proteoglycan monomers. Characterization of a sequence allowing the
RL synthesis of an oligonucleotide probe.";
RL FEBS Lett. 176:37-42(1984).
RN [6]
RP PARTIAL SEQUENCE.
RX MEDLINE=87005253; PubMed=3530809;
RN Perin J.-P., Bonnet F., Jolles P.;
RT "Structural relationship between link proteins and proteoglycan
RL monomers.";
RL FEBS Lett. 206:73-77(1986).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (by
CC similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P13608-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P13608-2; Sequence=VSP_003072;
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE C-TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND O-
CC LINKED (ABOUT 40) OLIGOSACCHARIDES.
CC -1- PTM: THE KERATAN SULFATE CONTENTS DIFFER CONSIDERABLY BETWEEN
CC ADULT AND FETAL BOVINE PROTEOGLYCANS.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 4 link domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; U76615; AAB38524.1; -; NOT_ANNOTATED_CDS.
CC EMBL; L07053; -; NOT_ANNOTATED_CDS.
CC PIR; A34234; A39808.
CC PIR; T42630; T42630.
CC HSSP; P08709; 1BF9.
CC InterPro; IPR002353; AntifreezeZell.
CC InterPro; IPR000152; Aex_hydroxyl.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF_Like.
CC InterPro; IPR007110; Ig_Like.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR003324; SGXSG.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00047; Ig; 1.
CC Pfam; PF00059; Lectin_C; 1.
CC Pfam; PF02339; SGXSG; 61.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 4.
CC PRINTS; PR01265; LINKMODULE.
CC PRINTS; PR00356; ANTIFREEZEII.
CC ProDom; PD000918; Link; 4.
CC SMART; SM00032; CCP; 1.
CC SMART; SM00034; CLECT; 1.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00445; LINK; 4.
CC PROSITE; PS00010; ASX_HYDROXYL; 1.
CC PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
CC PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01187; EGF_CA; 1.
CC PROSITE; PS00835; IG_LIKE; 1.
CC PROSITE; PS00290; IG_MHC; FALSE_NEG.
CC PROSITE; PS01241; LINK; 4.
CC Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
CC EGF-like domain; Calcium; Alternative splicing; Repeat;
CC Immunoglobulin domain.
CC SIGNAL 1 16 POTENTIAL.
CC CHAIN 17 2364 AGGRECAN CORE PROTEIN.
CC DOMAIN 25 147 IG-LIKE V-TYPE.
CC DOMAIN 170 247 LINK 1.
CC DOMAIN 268 349 LINK 2.
CC DOMAIN 504 581 LINK 3.
CC DOMAIN 602 683 LINK 4.
CC DOMAIN 774 907 LINK 5.
CC DOMAIN 1433 2112 23 X 6 AA APPROXIMATE TANDEM REPEATS OF
CC DOMAIN 2113 2149 E-[EK]-P-P-S.
CC DOMAIN 2114 2364 CS-2.
CC DOMAIN 2161 2276 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
CC DOMAIN 2280 2338 G3.
CC DISULFID 51 133 C-TYPE LECTIN.
CC DISULFID 175 246 BY SIMILARITY.
CC DISULFID 199 220 BY SIMILARITY.
CC DISULFID 273 348 BY SIMILARITY.

FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 509 580 BY SIMILARITY.
FT DISULFID 533 554 BY SIMILARITY.
FT DISULFID 607 682 BY SIMILARITY.
FT DISULFID 631 652 BY SIMILARITY.
FT DISULFID 2117 2128 BY SIMILARITY.
FT DISULFID 2182 2274 BY SIMILARITY.
FT DISULFID 2250 2266 BY SIMILARITY.
FT DISULFID 2281 2324 BY SIMILARITY.
FT DISULFID 2310 2337 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 2114 2150 Missing (in isoform 2).
FT FTid=VSP_003072.
SQ SEQUENCE 2364 AA; 246359 MW; 6FF83763420C3D4C CRC64;
Query Match 14.4%; Score 158.5; DB 1; Length 2364;
Best Local Similarity 26.8%; Pred. No. 7.7e-06;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;
QY 6 QPVCRGG--TORPCYKVIYFHTDTSRLNPEEAKACRRDGGQGLVSTESDEQKLEKFI 62
Db 2151 QKCEEGWTFQGHYR--HFPD---RATWDAESQCRKQSHLSIVTPEEQ-----EFV 2201
QY 63 ENLLPSDGF-WTGLRRREKQNSTACQDLYAWTDSISQFRNYYVDEP-----SCGSEV 117
Db 2202 NN---NAQDQWIGL-----NDKTIEGFRWSQSHLSQFENWRPNQPDNFATGEDC 2250
QY 118 CVMYHOPSPAGIGGYPYFQWNDRCNMKNFKYK-----DEKPAVPSREAGSETE 172
Db 2251 VVMIMHEKG-----EWNDVPCNYQLPFTCKRGTVACGEPVVEHARIFGQKKD 2298
RESULT 14
ID_PGCA HUMAN STANDARD; PRT; 2415 AA.
AC P16112; Q13650;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP) (Chondroitin sulfate proteoglycan core protein 1).
DE AC1 OR CSPG1.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Chondrocytes;
RX MEDLINE=91093289; PubMed=1985970;
RA Doegre K.J., Sasaki M., Kimura T., Yamada Y.;
RT "Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms.";
RT J. Biol. Chem. 266:1894-902 (1991).
RN [2]
RP SEQUENCE OF 1778-2415 FROM N.A. (ISOFORM 2).
RC TISSUE=Chondrocytes;
RA Dudhia J., Hardingham T.E.;
RL Submitted (JAN-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1936-2415 FROM N.A. (ISOFORM 1).
RX MEDLINE=89380154; PubMed=2789216;
RA Baldwin C.T., Reginato A.M., Prockop D.J.;
RT "A new epidermal growth factor-like domain in the human core protein for the large cartilage-specific proteoglycan. Evidence for alternative splicing of the domain.";
RT J. Biol. Chem. 264:15747-15750 (1989).

RP (4)
RC SEQUENCE OF 764-864 FROM N.A.
RX TISSUE=Blood;
RA MEDLINE=95128522; PubMed=7827755;
RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
RT "Length variation in the keratan sulfate domain of mammalian aggrecan.";
RL Matrix Biol. 14:323-328 (1994).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION.
CC -1- SUBUNIT: Interacts with FBLN1 (by similarity).
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (by similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P16112-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P16112-2; Sequence=VSP_003074;
CC Name=3;
CC IsoId=P16112-3; Sequence=VSP_003074, VSP_003075;
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 4 link domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M55172; AAA62824.1; -;
CC EMBL; J05062; AAA35726.1; -;
CC EMBL; X17406; CAA35463.1; -;
CC EMBL; S74659; AAC60643.2; -;
CC PIR; A39086; A39086.
CC HSSP; P98066; 1TSG.
CC Genew; HGNC:319; AGC1.
CC MTM; 155760; -;
CC GO; GO:0005204; F.chondroitin sulfate proteoglycan; TAS.
CC InterPro; IPR023353; Antifreeze1.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00047; ig; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF02339; SGXSG; 71.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 4.

DR PRINTS; PRO1265; LINKMODULE.
 DR PRINTS; PRO0356; ANTIFREEZEII.
 DR ProDom; PD000918; Link; 4.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 4.
 DR PROSITE; PS00615; C_TYPE LECTIN_1; 1.
 DR PROSITE; PS00041; C_TYPE LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; 1.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS00835; IG-LIKE; 1.
 DR PROSITE; PS00290; IG_MHC; 1.
 DR PROSITE; PS01241; LINK; 3.
 DR Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW EGF-like domain; Alternating splicing; Repeat; Immunoglobulin domain.
 FT SIGNAL 1 19 POTENTIAL
 FT CHAIN 20 2415 AGGREGAN CORE PROTEIN.
 FT DOMAIN 34 147 IG-LIKE V-TYPE.
 FT DOMAIN 170 247 LINK 1.
 FT DOMAIN 268 349 LINK 2.
 FT DOMAIN 495 572 LINK 3.
 FT DOMAIN 593 673 LINK 4.
 FT DOMAIN 2164 2199 EGF-LIKE.
 FT DOMAIN 2201 2327 C-TYPE LECTIN.
 FT DOMAIN 2331 2389 SUSHI.
 FT DOMAIN 48 141 G1-A.
 FT DOMAIN 152 247 G1-B.
 FT DOMAIN 253 349 G1-B'.
 FT DOMAIN 477 571 G2-B.
 FT DOMAIN 578 672 G2-B'.
 FT DOMAIN 676 848 KS.
 FT DOMAIN 772 844 12 X APPROXIMATE TANDEM REPEATS.
 FT DOMAIN 851 1497 29 X APPROXIMATE TANDEM REPEATS.
 FT DOMAIN 1498 2162 CS-1.
 FT DOMAIN 2163 2415 CS-2.
 FT DISULFID 51 133 BY SIMILARITY.
 FT DISULFID 175 246 BY SIMILARITY.
 FT DISULFID 199 220 BY SIMILARITY.
 FT DISULFID 273 348 BY SIMILARITY.
 FT DISULFID 297 318 BY SIMILARITY.
 FT DISULFID 500 571 BY SIMILARITY.
 FT DISULFID 524 545 BY SIMILARITY.
 FT DISULFID 598 672 BY SIMILARITY.
 FT DISULFID 621 642 BY SIMILARITY.
 FT DISULFID 2168 2178 BY SIMILARITY.
 FT DISULFID 2173 2187 BY SIMILARITY.
 FT DISULFID 2189 2198 BY SIMILARITY.
 FT DISULFID 2205 2216 BY SIMILARITY.
 FT DISULFID 2233 2325 BY SIMILARITY.
 FT DISULFID 2301 2317 BY SIMILARITY.
 FT DISULFID 2332 2375 BY SIMILARITY.
 FT DISULFID 2361 2388 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1898 1898 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 2163 2200 Missing (in isoform 2 and isoform 3).
 FT VARSPLIC 2330 2390 Missing (in isoform 3).
 FT CONFLICT 766 766 E -> A (IN REF. 4).
 FT CONFLICT 847 847 E -> V (IN REF. 4).
 FT CONFLICT 1928 1928 I -> V (IN REF. 2).
 FT CONFLICT 1964 1964 I -> V (IN REF. 2 AND 3).
 FT CONFLICT 2070 2070 P -> A (IN REF. 3).

FT CONFLICT 2391 2391 A -> P (IN REF. 2 AND 3).
 SQ SEQUENCE 2415 AA; 250191 MW; 1288937E1B98C6B6 CRC64;
 Query Match 14.4%; Score 158.5; DB 1; Length 2415;
 Best Local Similarity 29.1%; Pred. No. 7.9e-06;
 Matches 52; Conservative 22; Mismatches 62; Indels 43; Gaps 11;
 QY 6 QPVCRRG--TORPCYKVIYFHTDTSRLNFEKAEACRRDGGQLVSTSEDEQKLIKFI 62
 DB 2202 QEVCEGWNKYQHCYR--HFPD--RETWDARRCREQOQSHLSSIVTTEQ-----EFV 2252
 QY 63 ENLLPSDGF-WIGLRRREKQSNSTACQDLYAWTDGSIISQFRNYYVDFPS---CGSEVC 118
 DB 2253 NN---NAQDYQWIGL-----NDRITGDPKWDGHPQFENRPNQPDNFFAAGEDC 2301
 QY 119 VVM-YHQPSAPAGIGGYPQWDDRCNWKWFIKYS-----DEKPAVPSPREAREGETE 172
 DB 2302 VVMWHEKG-----EMNDVPCNVHLPTCKKGTACGPPVVEHARTFGOKKD 2349

RESULT 15
 PGCA CANFA
 ID PGCA CANFA STANDARD; PRT; 2333 AA.
 AC Q28343; Q28310;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
 DE GN AGC1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Glant T.T., Adams M.E., Kwok S.X.F., Huang D., Fuloop C.;
 RT "Complete coding sequence and deduced amino acid sequence of aggrecan of canine cartilage".
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 774-833 FROM N.A.
 RC TISSUE=Cartilage;
 RX MEDLINE=95128522; PubMed=7827755;
 RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
 RT "Length variation in the keratan sulfate domain of mammalian aggrecan".
 RL Matrix Biol. 14:323-328(1994).
 RN [3]
 RP SEQUENCE OF 1830-2333 FROM N.A.
 RA Adams M.E., Kwok S.X.F., Huang D., Glant T.T., Fuloop C.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 2082-2118 FROM N.A.
 RC TISSUE=Cartilage;
 RX MEDLINE=93352525; PubMed=8349621;
 RA Fuloop C., Walcz E., Valyon M., Glant T.T.;
 RT "Expression of alternatively spliced epidermal growth factor-like domains in aggrecans of different species. Evidence for a novel module".
 RL J. Biol. Chem. 268:17377-17383(1993).
 CC - FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC - SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
 CC - DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)

CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
 CC AND G3.
 CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
 CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 4 link domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: Contains 1 EGF-like domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U65989; AAB06238.2; -.
 CC EMBL; S74562; AAC60527.1; -.
 CC EMBL; L07054; -; NOT_ANNOTATED_CDS.
 CC PIR; I46998; I46998.
 CC HSRP; P08709; IBF9.
 CC InterPro; IPR000152; Asx hydroxyl.
 CC InterPro; IPR00742; EGF 2.
 CC InterPro; IPR001881; EGF Ca.
 CC InterPro; IPR006209; EGF-like.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR003006; Ig_MHC.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR005338; Link.
 CC InterPro; IPR003324; SGXSG.
 CC InterPro; IPR000436; Sushi_SCR_CCP.
 CC Pfam; PF00008; EGF; 1.
 CC Pfam; PF00047; Ig; 1.
 CC Pfam; PF00059; lectin_c; 1.
 CC Pfam; PF02339; SGXSG; 66.
 CC Pfam; PF00084; sushi; 1.
 CC Pfam; PF00193; Xlink; 4.
 CC PRINTS; PR01365; LINKMODULE.
 CC PRINTS; PR00356; ANTIFREEZEII.
 CC ProDom; PD000918; Link; 4.
 CC SMART; SM00032; CCP; 1.
 CC SMART; SM00034; CLECT; 1.
 CC SMART; SM00179; EGF CA; 1.
 CC SMART; SM00445; LINK; 4.
 CC PROSITE; PS00010; ASX HYDROXYL; 1.
 CC PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 CC PROSITE; PS50041; C-TYPE LECTIN_2; 1.
 CC PROSITE; PS00022; EGF 1; 1.
 CC PROSITE; PS01187; EGF CA; 1.
 CC PROSITE; PS50835; IG LIKE; 1.
 CC PROSITE; PS00290; IG_MHC; 1.
 CC PROSITE; PS01241; LINK; 4.
 CC KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 CC EGF-like domain; Repeat; Immunoglobulin domain.
 CC SIGNAL 1 16
 CC CHAIN 17 2333
 CC DOMAIN 34 147
 CC LINK 1.
 CC DOMAIN 170 247
 CC LINK 2.
 CC DOMAIN 268 349
 CC LINK 3.
 CC DOMAIN 513 590
 CC LINK 4.
 CC DOMAIN 611 692
 CC LINK 4.
 CC DOMAIN 2081 2117
 CC EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
 CC C-TYPE LECTIN.
 CC DOMAIN 2130 2245
 CC SUSHI.
 CC DOMAIN 2249 2307
 CC GI-A.
 CC DOMAIN 48 140
 CC GI-B.
 CC DOMAIN 152 247
 CC GI-B'.
 CC DOMAIN 253 349
 CC G2-B'.
 CC DOMAIN 495 589
 CC G2-B'.
 CC DOMAIN 596 691
 CC KS.
 CC DOMAIN 694 816

FT DOMAIN 819 1394
 FT DOMAIN 1395 2079
 FT DOMAIN 2080 2333
 FT DISULFID 51 133
 FT DISULFID 175 246
 FT DISULFID 199 220
 FT DISULFID 273 348
 FT DISULFID 297 318
 FT DISULFID 518 589
 FT DISULFID 542 563
 FT DISULFID 616 691
 FT DISULFID 640 661
 FT DISULFID 2085 2096
 FT DISULFID 2090 2105
 FT DISULFID 2107 2116
 FT DISULFID 2123 2134
 FT DISULFID 2151 2243
 FT DISULFID 2219 2235
 FT DISULFID 2250 2293
 FT CARBOHYD 2279 2306
 FT CARBOHYD 126 126
 FT CARBOHYD 239 239
 FT CARBOHYD 333 333
 FT CARBOHYD 387 387
 FT CARBOHYD 444 444
 FT CARBOHYD 620 620
 FT CARBOHYD 676 676
 FT CARBOHYD 747 747
 SQ SEQUENCE 2333 AA; 240573 MW; 8B9ED78F3508B596 CRC64;
 Query Match 14.1%; Score 155.5; DB 1; Length 2333;
 Best Local Similarity 28.5%; Pred. No. 1.4e-05;
 Matches 51; Conservative 24; Mismatches 61; Indels 43; Gaps 11;
 QY 6 QPVCRGG---TQRPCKVIYFHDTSRRNLNPFBEAKEACRRDGGQLVSISEDEKLIKRFI 62
 Db 2120 QELCEEGWTKFGHCYR--YFPD---RESWDAESRCRAQQSHLSIVTPEQ---BVF 2170
 QY 63 ENLLPSDGDFF-WIGLRREEKQSNSTACQDIYANTDGSISQFRNRYVDEPS---CGSEVC 118
 Db 2171 NN---NAQDYQWIGL-----NDRITEGDFRWSGHSLOFENRPNQPNFFVSGEDC 2219
 QY 119 VVM-YHOPSAFAGIGGPPYMFQWDDRCNMKNFKICKYS---DEKPAVPSREAGEETE 172
 Db 2220 VVMWHEKG-----EWDVPCNYLPLFTCKKGTACGDPVVEHARTFGQKKD 2267

Search completed: December 22, 2003, 16:10:59
 Job time : 7.46605 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:06:34 ; Search time 9.86924 Seconds
(without alignments)
1978.090 Million cell updates/sec

Title: US-09-887-855-2_COPY_25_227

Perfect score: 1100
Sequence: 1 RLISGQVCRGQTQPCYKV.....EEDAKTKFSREAAALNLAY 203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: Piri:*

2: Piri:*

3: Piri:*

4: Piri:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 182 | 16.5 | 1456 | 1 A36563 | mannose receptor p |
| 2 | 178.5 | 16.2 | 1455 | 1 A48925 | mannose receptor p |
| 3 | 177 | 16.1 | 1268 | 2 S52781 | neurocan - mouse |
| 4 | 174.5 | 15.9 | 1643 | 2 T14274 | versican precursor |
| 5 | 174.5 | 15.9 | 3381 | 2 T42389 | versican precursor |
| 6 | 174 | 15.8 | 1257 | 2 S28764 | neurocan precursor |
| 7 | 174 | 15.8 | 2397 | 1 A55335 | versican precursor |
| 8 | 174 | 15.8 | 2409 | 1 A60979 | versican precursor |
| 9 | 171 | 15.5 | 3562 | 2 A47171 | chondroitin sulfat |
| 10 | 170.5 | 15.5 | 1479 | 2 T42710 | mannose receptor, |
| 11 | 158.5 | 14.4 | 1340 | 2 A39808 | proteoglycan core |
| 12 | 158.5 | 14.4 | 2327 | 2 T42830 | aggreacan - bovine |
| 13 | 158.5 | 14.4 | 2415 | 1 A39086 | aggreacan precursor |
| 14 | 154.5 | 14.0 | 612 | 2 B42755 | E-selectin precurs |
| 15 | 153.5 | 14.0 | 2124 | 2 A28452 | proteoglycan core |
| 16 | 152 | 13.8 | 912 | 2 A54423 | hypothetical prote |
| 17 | 149.5 | 13.6 | 459 | 2 T24425 | hypothetical prote |
| 18 | 149 | 13.5 | 321 | 1 LNHUP6 | IGF Fc receptor II |
| 19 | 148.5 | 13.5 | 330 | 2 T46256 | protein F52E1.2 (i |
| 20 | 148 | 13.5 | 253 | 2 E89130 | aggreacan precursor |
| 21 | 147.5 | 13.4 | 2132 | 1 A55182 | aggreacan precursor |
| 22 | 147 | 13.4 | 883 | 2 S57653 | aggreacan precursor |
| 23 | 146.5 | 13.3 | 162 | 1 LNR01 | lectin BRA3-1 prec |
| 24 | 146 | 13.3 | 2109 | 1 I50421 | aggreacan precursor |
| 25 | 145.5 | 13.2 | 742 | 2 JCT595 | scavenger receptor |
| 26 | 145 | 13.2 | 883 | 2 S49126 | brevican precursor |
| 27 | 144.5 | 13.1 | 173 | 2 S10548 | lectin - barnacle |
| 28 | 144.5 | 13.1 | 372 | 2 S23936 | L-selectin precurs |
| 29 | 144.5 | 13.1 | 404 | 2 A46274 | HIV gp120-binding |

QY 18 YKVIYFHDTRRLNFEAKEACRRDGGQLVSIETSEDEQKLIETKFIENLLPSDGFWIGLR 77

Query Match 16.5%; Score 182; DB 1; Length 1456;
Best Local Similarity 25.5%; Pred. No. 8.2e-08;
Matches 52; Conservative 37; Mismatches 61; Indels 54; Gaps 8;

ALIGNMENTS

RESULT 1

A36563

mannose receptor precursor - human

C/Species: Homo sapiens (man)

C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999

C/Accession: A36563; A60926; A44255; B44255; C44255; D44255; E44255; F44255; G44255; H44255; I44255; J44255; K44255; L44255; M44255; N44255; O44255; P44255; Q44255; R44255; S44255; T44255; U44255; V44255; W44255; X44255; Y44255; Z44255

R/Taylor, M.E.; Conary, J.T.; Lennartz, M.R.; Stahl, P.D.; Drickamer, K.

J. Biol. Chem. 265, 12156-12162, 1990

A/Title: Primary structure of the mannose receptor contains multiple motifs resembling

A/Reference number: A36563; MUID:90324192; PMID:2373685

A/Accession: A36563

A/Molecule type: mRNA

A/Residues: 1-1456 <TAY>

A/Cross-references: GB:J05550; NID:9188675; PIDN:AA59868.1; PID:9188676

A/Note: parts of this sequence, including the amino end of the mature protein, were cor

R/Ekowitz, R.A.B.; Sastry, K.; Bailly, P.; Warner, A.

J. Exp. Med. 172, 1785-1794, 1990

A/Title: Molecular characterization of the human macrophage mannose receptor: demonst

A/Reference number: A60926; MUID:91079783; PMID:2258707

A/Accession: A60926

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-1333, 'T', 1335-1456 <EZE>

A/Cross-references: GB:X55635

A/Note: translation of the nucleotide sequence is incomplete

A/Note: in the authors' translation additional residues Pro-Glu-Ile are shown after 49;

R/Kim, S.J.; Ruiz, N.; Bezouska, K.; Drickamer, K.

Genomics 14, 721-727, 1992

A/Title: Organization of the gene encoding the human macrophage mannose receptor (MRC1)

A/Reference number: A44255; MUID:91052405; PMID:1294118

A/Accession: A44255

A/Status: Preliminary; nucleic acid sequence not shown; not compared with conceptual t

A/Molecule type: DNA

A/Residues: 155-233, 'KSA', 238-283, 346-428, 492-569, 631-714, 716-719, 783-820, 'N', 822-865,

A/Note: sequence extracted from NCBI backbone (NCBIP:118415, NCBIP:118421, NCBIP:118442)

C/Genetics:

A/Gene: GDB:MRC1

A/Cross-references: GDB:133759; OMIM:153618

A/Map position: 10p13-10p13

C/Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II ;

F/1-18/Domain: duplication; lectin; tandem repeat; transmembrane protein

F/168-209/Domain: signal sequence #status predicted <SIG>

F/223-340/Domain: fibronectin type II repeat homology <2F1>

F/362-486/Domain: C-type lectin homology <LCH1>

F/945-1079/Domain: C-type lectin homology <LCH2>

```
Db 807 YKDQYVFSKEKTMNDNARACKRNGDLVSIQSESEKFLWKYV-NRNDQAQSAFYIGLL 865
QY 78 RREEKQSNSTACQDLVATDGSISQFRNMYVDPSGCS--EVCVMYHQPSPAGIGGYP 135
Db 866 ISLDKK-----FAMWMDGSKVDYVSMATGEPFNANEDENCVTMY-----SNSGF---- 908
QY 136 MFQWDDRCNMKNFKICKYSDK-----PAVPSREAGEETEELTPVLPEETQ----- 184
Db 909 ---WNDINGCYPNAFICQRHNSINATTMP-----TMPSPVSCCKGWNFYSN 954
QY 185 -----EDAKKTFKESREAAAL 199
Db 955 KCFKIFGFMEERKNMQEAKACI 978

RESULT 2
A08925
mannose receptor precursor, macrophage - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A48925; S21320; PC2245
R:Harris, N.; Super, M.; Rits, M.; Chang, G.; Ezekowitz, R.A.
Blood 80, 2363-2373, 1992
A:Title: Characterization of the murine macrophage mannose receptor: demonstration that
A:Reference number: A48925; MUID:93043353; PMID:1421407
A:Accession: A48925
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1455 <HA>
A:Experimental source: peritoneal macrophage
A>Note: sequence extracted from NCBI backbone (NCBIPI:118733)
R:Harris, N.; Super, M.; Rits, M.; Chang, G.; Ezekowitz, R.B.
submitted to the EMBL Data Library, April 1992
A:Description: Characterization of the murine macrophage mannose receptor: Demonstration
on.
A:Reference number: S21320
A:Accession: S21320
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-302, 'W', 303-1117, 'E', 1118-1455 <HA2>
A:Cross-references: EMBL:Z11974; NID:G52997; PIDN:CAA78028.1; PID:G52998
R:Harris, N.; Peters, L.L.; Eichler, E.M.; Rits, M.; Raspberry, D.; Eichbaum, Q.G.; Super
Biochem. Biophys. Res. Commun. 198, 682-692, 1994
A:Title: The exon-intron structure and chromosomal localization of the mouse macrophage
A:Reference number: PC2245; MUID:94128116; PMID:8297379
A:Accession: PC2245
A:Molecule type: mRNA
A:Residues: 35-105 <HA3>
C:Genetics:
A:Gene: Mrcl
A:Map position: 2
C:Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II re
C:Keywords: membrane protein; receptor
F:168-209/Domain: fibronectin type II repeat homology <2F9>
F:361-485/Domain: C-type lectin homology <LCHI>
F:943-1077/Domain: C-type lectin homology <LCHI2>

Query Match 16.2%; Score 178.5; DB 1; Length 1455;
Best Local Similarity 25.9%; Pred. No. 1.7e-07;
Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;

QY 18 YKVIYFHTSRRLNFEAKEACRRDGGQGVLSIESDEQKLIKFIENLLPSDGFWIGLR 77
Db 806 YKDQYVFSKEKTMNDNARRCKNGFDLATIKSEKFLWKYI-NKNGGQSPYFIGNL 864
QY 78 RREEKQSNSTACQDLVATDGSISQFRNMYVDPSGCS--EVCVMYHQPSPAGIGGYP 135
Db 865 ISMDKK-----FIMWMDGSKVDVAVATGEPFNANDENCVTMY-----TNSGF---- 907
QY 136 MFQWDDRCNMKNFKICK---YSDEKPAVPSREAGEETEELTPVLPEETQ----- 184
Db 908 ---WNDINGCYPNPFICQRHNSINATAMP-----TTPTPGCKEGWHLYKNK 953
```

```
QY 185 -----EDAKKTFKESREAAALNL 201
Db 954 CTKIFGFANEEKSKSQDARQACKGL 978

RESULT 3
S52781
neurocan - mouse
C:Species: Mus musculus (house mouse)
C>Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 04-Feb-2000
C:Accession: S52781
R:Rauch, U.; Foreberg, N.; Kulbe, G.; Arnold-Ammer, I.; Faessler, R.
submitted to the EMBL Data Library, February 1995
A:Description: Amino acid sequence of mouse neurocan and brevican and their different
A:Reference number: S52781
A:Accession: S52781
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1268 <RAU>
A:Cross-references: EMBL:X84727; NID:G758629; PIDN:CAA59216.1; PID:G758630
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-253/Domain: link protein repeat homology <LNK2>
F:964-995/Domain: EGF homology <EGF>
F:1040-1160/Domain: C-type lectin homology <LCH>
F:1167-1223/Domain: complement factor H repeat homology <FHD>

Query Match 16.1%; Score 177; DB 2; Length 1268;
Best Local Similarity 31.5%; Pred. No. 1.9e-07;
Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;

QY 14 QRCPKVIYFHTSRRLNFEAKEACRRDGGQGVLSIESDEQKLIKFIENLLPSDGDPM 73
Db 1048 QGHCYR--YF---AHRRAWEDAECDRRRAGHLTSVHSPEEHKFINSF-----GHENSW 1096
QY 74 IGLRREEKQSNSTACQDLVATDGSISQFRNMYVDPS---CGSEVCVMYHQPSPAG 130
Db 1097 IGLNDRITVERD-----FQWTDNTGLQYENREKQPDNFFAGGDCVVMVAHESG--- 1145
QY 131 ICGPYMFQWDDRCNMKNFKICK 153
Db 1146 -----RMNDVPCNYNLPYVCK 1161

RESULT 4
T14274
versican precursor, splice form V2 - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000
C:Accession: T14274
R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
J. Biol. Chem. 273, 15758-15764, 1998
A:Title: Versican V2 is a major extracellular matrix component of the mature bovine bra
A:Reference number: Z17954; MUID:98288320; PMID:9624174
A:Accession: T14274
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1643 <SCH>
A:Cross-references: EMBL:AF060458; NID:G32533303; PID:G32533304; PIDN:AAC24360.1
A:Experimental source: brain
C:Keywords: glycoprotein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-1643/Product: versican, splice form V2 #status predicted <MAT>
F:57,331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (asn) (covalent) #s

Query Match 15.9%; Score 174.5; DB 2; Length 1643;
Best Local Similarity 25.5%; Pred. No. 4.4e-07;
Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;

QY 14 QRCPKVIYFHTSRRLNFEAKEACRRDGGQGVLSIESDEQKLIKFIENLLPSDGDPM 72
Db 1424 QGQCYK--YF---AHRRTWDAARECRQLGAHLTSILSHEEQMFVNRV-----GHDYQ 1471
```

F:23-1257/Product: neurocan #status predicted <MAT>
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-355/Domain: link protein repeat homology <LNK2>
F:364-366/Region: cell attachment (R-G-D) motif
F:953-984/Domain: EGF homology <EGF>
F:1029-1149/Domain: C-type lectin homology <LCH>
F:1126-1212/Domain: complement factor H repeat homology <FHD>
F:121,339,737,967,1164/Binding site: carbonylrate (Asn) (covalent) #status predicted
F:377,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F:1944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

Query Match 15.8%; Score 174; DB 2; Length 1257;
Best Local Similarity 30.8%; Pred. No. 3.5e-07;
Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;

Qy 14 QRCYKVIYFHDTSRLNFEAEACRRDGGQLVSISEDEQKLEFIENLLPSDGF 73
Db 1037 QGCHYR--YF---AHRRAWEDAECDRCRRAGHLTSVHSPEHKPINSF-----CHENSW 1085

Qy 74 IGLRRREKQSNSTACQDLVWTDGSIQFRNMYVDEPS-----CGSEVCVMYHQPSAP 130
Db 1086 IGLNDRTVRD-----FQMTDNTGLQYENWREKQPNFNFAGGDCVVMVAHENG--- 1134

Qy 131 IGGPYMFQWDDRCNMKNPFICK 153
Db 1135 -----RWNDVPCNYNLPVYCK 1150

RESULT 7
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-2397 <RES>
A:Cross-references: GB:D16263; NID:G862460; PIDN:BAA03796.1; PID:G862461
C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology;
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-1254/Domain: versican #status predicted <MAT>
F:167-244/Domain: link protein repeat homology <LNK1>
F:265-346/Domain: link protein repeat homology <LNK2>
F:2095-2126/Domain: EGF homology <EG1>
F:2133-2164/Domain: EGF homology <EG2>
F:2171-2291/Domain: C-type lectin homology <LCH>
F:2298-2354/Domain: complement factor H repeat homology <FHD>

Query Match 15.8%; Score 174; DB 1; Length 2397;
Best Local Similarity 28.5%; Pred. No. 7.5e-07;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

Qy 14 QRCYKVIYFHDTSRLNFEAEACRRDGGQLVSISEDEQKLEFIENLLPSDGF- 72
Db 2179 QGCHYR--YF---AHRRTWDAERCLQAGHLTSILSHEEQMFVNRV-----GHDYQ 2226

Qy 73 WIGLRRREKQSNSTACQDLVWTDGSIQFRNMYVDEP-----SCGSEVCVMYHQPSAP 128
Db 2227 WIGL-----NDKMFEDFRWTDGSLQYENWRNRPQSPFSAGDCVVIWHENG-- 2276

Qy 129 AGIGPYMFQWDDRCNMKNPFICKYS-----DEKPAVPSREARGE 169
Db 2277 -----QWNDVPCNYHLTYTCKGTVACGQPPVVENAKTFGK 2312

1472 WIGL-----NDKMFEDFRWTDGSLQYENWRNRPQSPFSAGDCVVIWHENG-- 1521

Qy 129 AGIGPYMFQWDDRCNMKNPFICKYS-----DEKPAVPSREARGE----- 169
Db 1522 -----QWNDVPCNYHLTYTCKGTVACGQPPVVENAKTFGKMPRYEINSLIRYHC 1572

Qy 170 -----ETELTT-----PVL-----PEETQEDAKTFKESREAA 200
Db 1573 KGFQIRHLPTIRCLNGRWAMPKITCLNPSAYQRTYSKYFKNSSAKDN 1623

RESULT 5
T42389
N:Alternate names: versican precursor, splice form V0 - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 05-May-2000
C:Accession: T42389
R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
J. Biol. Chem. 273, 15758-15764, 1998
A:Title: Versican V2 is a major extracellular matrix component of the mature bovine brain
A:Reference number: Z17954; MUID:98288320; PMID:9624174
A:Accession: T42389
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3381 <SCH>
A:Cross-references: EMBL:AF060456; NID:G3253299; PID:G3253300; PIDN:AAC24358.1
C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lectin
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-3381/Product: versican, splice form V0 #status predicted <MAT>
F:57,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2356,26

Query Match 15.9%; Score 174.5; DB 2; Length 3381;
Best Local Similarity 25.5%; Pred. No. 1e-06;
Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;

Qy 14 QRCYKVIYFHDTSRLNFEAEACRRDGGQLVSISEDEQKLEFIENLLPSDGF- 72
Db 3162 QGCHYR--YF---AHRRTWDAERCLQAGHLTSILSHEEQMFVNRV-----GHDYQ 3209

Qy 73 WIGLRRREKQSNSTACQDLVWTDGSIQFRNMYVDEP-----SCGSEVCVMYHQPSAP 128
Db 3210 WIGL-----NDKMFEDFRWTDGSLQYENWRNRPQSPFSAGDCVVIWHENG-- 3259

Qy 129 AGIGPYMFQWDDRCNMKNPFICKYS-----DEKPAVPSREARGE----- 169
Db 3260 -----QWNDVPCNYHLTYTCKGTVACGQPPVVENAKTFGKMPRYEINSLIRYHC 3310

Qy 170 -----ETELTT-----PVL-----PEETQEDAKTFKESREAA 200
Db 3311 KGFQIRHLPTIRCLNGRWAMPKITCLNPSAYQRTYSKYFKNSSAKDN 3361

RESULT 6
S28764
neurocan precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
C:Accession: S28764
R:Rauch, U.; Karthikayan, L.; Maurel, P.; Margolis, R.U.; Margolis, R.K.
J. Biol. Chem. 267, 19536-19547, 1992
A:Title: Cloning and primary structure of neurocan, a developmentally regulated, aggregated
A:Reference number: S28764; MUID:92406907; PMID:1326557
A:Accession: S28764
A:Molecule type: mRNA
A:Residues: 1-1257 <RAU>
A:Cross-references: EMBL:M97161; NID:G205649; PIDN:AAC37679.1; PID:G205650
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EGF
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein

| P.1-22/nmain. airmal samianco #status needed -GTC-

RESULT 8
A60979
versican precursor - human
N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan
C:Contains: glial hyaluronate-binding protein
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 19-Jan-2001
C:Accession: S06014; S43921; A60979; A30358; A29348; A45131; I54179
R:Zimmermann, D.R.; Ruoslahti, E.
EMBO J. 8, 2375-2381, 1989
A:Title: Multiple domains of the large fibroblast proteoglycan, versican.
A:Reference number: S06014; MUID:90059882; PMID:2583089
A:Accession: S06014
A:Molecule type: mRNA
A:Residues: 1-2409 <ZIN>
A:Cross-references: GB:X15998; NID:g37662; PIDN:CAA34128.1; PID:g37663
R:Yao, L.Y.; Moody, C.; Schoenher, E.; Wight, T.N.; Sandell, L.J.
Matrix Biol. 14, 213-225, 1994
A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by
A:Reference number: S43921; MUID:95005762; PMID:7921538
A:Accession: S43921
A:Molecule type: mRNA
A:Residues: 208-440;1094-1385;1910-2246 <YAO>
R:Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.
Brain Res. Bull. 22, 67-70, 1989
A:Title: Structural similarity of hyaluronate binding proteins in brain and cartilage.
A:Reference number: A60979; MUID:89229983; PMID:2469524
A:Accession: A60979
A:Molecule type: protein
A:Residues: 171-210;289-303 <BIG>
R:Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.
J. Biol. Chem. 264, 5981-5987, 1989
A:Title: Isolation and partial characterization of a glial hyaluronate-binding protein.
A:Reference number: A30358; MUID:89174663; PMID:2466833
A:Accession: A30358
A:Molecule type: protein
A:Residues: 24-50;80-87,'D',89-119;128-155;167-218;229-259,'IR';261-268;277-283,'G';285-
R:Krusius, T.; Gehlsen, K.R.; Ruoslahti, E.
J. Biol. Chem. 262, 13120-13125, 1987
A:Title: A fibroblast chondroitin sulfate proteoglycan core protein contains lectin-like
A:Reference number: A29348; MUID:88007514; PMID:2820964
A:Accession: A29348
A:Molecule type: mRNA
A:Residues: 1725,'V',1727-2409 <KRU>
A:Cross-references: GB:J02814
R:Perides, G.; Rahemtulla, F.; Lane, W.S.; Asher, R.A.; Bignami, A.
J. Biol. Chem. 267, 23883-23887, 1992
A:Title: Isolation of a large aggregating proteoglycan from human brain.
A:Reference number: A45131; MUID:93054750; PMID:1429726
A:Contents: brain
A:Accession: A45131
A:Molecule type: protein
A:Residues: 21-22,'X',24-37 <PE2>
A:Experimental source: brain
A:Note: Sequence extracted from NCBI backbone (NCBIP:118884)
R:Iozzo, R.V.; Naso, M.F.; Cannizzaro, L.A.; Wasmuth, J.J.; McPherson, J.D.
Genomics 14, 845-851, 1992
A:Title: Mapping of the versican proteoglycan gene (CSPG2) to the long arm of human chr
A:Reference number: I54179; MUID:93122792; PMID:1478664
A:Accession: I54179
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 251-347 <RES>
A:Cross-references: GB:S52488; NID:g263313; PIDN:AAB24878.1; PID:g263314
C:Genetics:
A:Gene: GDB:CSPG2
A:Cross-references: GDB:127873; OMIM:118661
A:Map position: 5q12-5q14
A:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EG
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-2409/Product: proteoglycan 24K core protein #status predicted <MAT>
F:167-244/Domain: link protein repeat homology <LNK1>

F:265-346/Domain: link protein repeat homology <LNK2>
F:559-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>
F:2106-2137/Domain: EGF homology <EG1>
F:2142-2175/Domain: EGF homology <EG2>
F:2182-2302/Domain: C-type lectin homology <LCH>
F:2309-2365/Domain: complement factor H repeat homology <FHD>
Query Match 15.8%; Score 174; DB 1; Length 2409;
Best Local Similarity 28.5%; Pred. No. 7.5e-07;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
QY 14 ORPCYKVIYFHDTSRRLNFEAEKACRRDGGOLVSISEDEQKLEKFIENLLPSDGF- 72
DB 2190 QGQCYK--YF---AHRRTWDAARECRLLQGAHLTSLSHEEQMFVNRV-----GHDYQ 2237
QY 73 WGLRRREKQSNSTACQDLYAWTDGSIQFRWYVDEP-----SCGSEVCVVMYHQSAP 128
DB 2238 WIGL-----NDKMFEDHFRWTDGSLQYENWRPNQDPSFSGEDCVIWHENG-- 2287
QY 129 AGIGGYPYFQWDDRCNMKNFICKYS---DEKPAVPSREAGE 169
DB 2288 -----QWNVPCNVHLTYTCKGTVACGQPPVVENAKTFGK 2323
RESULT 9
A47171
chondroitin sulfate proteoglycan PG-M core protein - chicken
C:Species: Gallus gallus (chicken)
C:Date: 21-Sep-1993 #sequence revision 18-Nov-1994 #text change 21-Jul-2000
C:Accession: A47171
R:Shinomura, T.; Nishida, Y.; Ito, K.; Kimata, K.
J. Biol. Chem. 268, 14461-14469, 1993
A:Title: CDNA cloning of PG-M, a large chondroitin sulfate proteoglycan expressed durin
A:Reference number: A47171; MUID:93300846; PMID:8314802
A:Accession: A47171
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-3562 <SHI>
A:Cross-references: GB:D13542; NID:g391643; PIDN:BA02742.1; PID:g391644
A:Experimental source: stage 22-23 developing limb buds
A:Note: sequence extracted from NCBI backbone (NCBIN:134456, NCBIP:134457)
C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lecti
F:166-243/Domain: link protein repeat homology <LNK1>
F:264-345/Domain: link protein repeat homology <LNK2>
F:3258-3289/Domain: EGF homology <EGF1>
F:3296-3327/Domain: EGF homology <EGF>
F:3334-3454/Domain: C-type lectin homology <LCH>
F:3461-3517/Domain: complement factor H repeat homology <FHD>
Query Match 15.5%; Score 171; DB 2; Length 3562;
Best Local Similarity 28.5%; Pred. No. 2.2e-06;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
QY 14 ORPCYKVIYFHDTSRRLNFEAEKACRRDGGOLVSISEDEQKLEKFIENLLPSDGF- 72
DB 3342 QGQCYK--YF---AHRRTWDAARECRLLQGAHLTSLSHEEQVFNRI-----GHDYQ 3389
QY 73 WGLRRREKQSNSTACQDLYAWTDGSIQFRWYVDEP-----SCGSEVCVVMYHQSAP 128
DB 3390 WIGL-----NDKMFEDHFRWTDGSLQYENWRPNQDPSFSGEDCVIWHENG-- 3439
QY 129 AGIGGYPYFQWDDRCNMKNFICKYS---DEKPAVPSREAGE 169
DB 3440 -----QWNVPCNVHLTYTCKGTVACGQPPVVENAKTFGK 3475
RESULT 10
T42710
mannose receptor, macrophage - mouse
N:Alternate names: lambda lectin; phospholipase A2 receptor
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text change 09-Jun-2000

```

C;Accession: T42710
R;Yu, K.; Yuan, J.; Lasky, L.A.
J. Biol. Chem. 271, 21323-21330, 1996
A;Title: Characterization of a novel member of the macrophage mannose receptor type C le
A;Reference number: Z22235; MUID:9635501; PMID:8702911
A;Accession: T42710
A;Status: preliminary; translated from GB/EMBL/DBDB
A;Molecule type: mRNA
A;Residues: 1-1479 <WUK>
A;Cross-references: EMBL:U56734; NID:g1336073; PID:g1336074; PIDN:AACS2729.1
C;Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II re
C;Keywords: membrane protein; receptor
F;186-227/Domain: fibronectin type II repeat homology <2FR>

Query Match      15.5%; Score 170.5; DB 2; Length 1479;
Best Local Similarity 31.4%; Pred. No. 8.7e-07;
Matches 48; Conservative 22; Mismatches 48; Indels 35; Gaps 7;

Qy      4  SGQVPCRGGTQPCYKVIYFHTDSRRLLNFEEAKEACRDGGQLVSTSEDEQKLIEKFIE 63
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      384 SWQPF-----QGHCYRL-----QAEKKSWQESRRACLRGGGDLTSHMALEFIFTKQIK 433

Qy      64  NLLPSDGDFFWIGLRRREEKOSNSTACQDLVATDGTSGISQFRNNWYVDEPS---CGSVCVV 120
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      434 QEVE---ELWIGL-----NDLKLQMFWSGSLVSTFTHWHPPEPNNFRDSDLCVT 482

Qy      121  MYHQPSAPAGIGPYFMQWDDRCNMKNFICK 153
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      483 IW----GPEG-----RWNDSPCNOSLPISCK 504

```

RESULT 11

A39808
proteoglycan core protein, cartilage - bovine (fragments)
N;Alternate names: aggrecan; aggregating cartilage proteoglycan
C;Species: Bos primigenius taurus (cattle)
C;Date: 20-Mar-1992 #sequence revision 23-Mar-1995 #text change 13-Aug-1999
C;Accession: A34234; A27752; A39808; A27751; E29164; B27751; C27751; E27751; P27751; F27751; R;Antonsson, P.; Heinegard, D.; Oldberg, A.
J. Biol. Chem. 264, 16170-16173, 1989
A;Title: The keratan sulfate-enriched region of bovine cartilage proteoglycan consists of
A;Reference number: A34234; MUID:89380219; PMID:2528543
A;Accession: A34234
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 128-621 <ANT>
A;Cross-references: GB:J05028
R;Oldberg, A.; Antonsson, P.; Heinegard, D.
Biochem. J. 243, 255-259, 1987
A;Title: The partial amino acid sequence of bovine cartilage proteoglycan, deduced from a
A;Reference number: A27752; MUID:87270630; PMID:3111460
A;Accession: A27752
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 622-1340 <OLD>
R;Sandy, J.D.; Boynton, R.E.; Flannery, C.R.
J. Biol. Chem. 266, 8198-8205, 1991
A;Title: Analysis of the catabolism of aggrecan in cartilage explants by quantitation of
A;Reference number: A39808; MUID:91217051; PMID:2022637
A;Accession: A39808
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-28;59-82;131-137, 'QSET', 142-149;196-207;226-249;1137-1143;1252-1267;1274-1
R;Perin, J.P.; Bonnet, F.; Jolles, P.
FEBS Lett. 206, 73-77, 1986
A;Title: Structural relationship between link proteins and proteoglycan monomers.
A;Reference number: A27751; MUID:87005253; PMID:3530809
A;Accession: A27751
A;Molecule type: protein
A;Residues: 29-58;74-130;174-175, 'A', 177-204;208-225 <PER>
R;Perin, J.P.; Bonnet, F.; Jolles, J.; Jolles, P.
FEBS Lett. 176, 37-42, 1984
A;Title: Sequence data concerning the protein core of the cartilage proteoglycan monomer.

```

A:Reference number: A91327; MUID:85027710; PMID:6489519
A:Accession: E29164
A:Molecule type: protein
A:Residues: 1230-1249 <PR2>
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
C:Keywords: glycoprotein
F:8-28/Domain: link protein repeat homology (fragment) <LNK1>
F:29-58/Domain: link protein repeat homology (fragment) <LNK2>
F:80-146/Domain: link protein repeat homology (fragments) <LNK3>
F:167-248/Domain: link protein repeat homology <LNK4>
F:1130-1250/Domain: C-type lectin homology <LCH>
F:1257-1313/Domain: complement factor H repeat homology <FHD>

Query Match          14.4%; Score 158.5; DB 2; Length 1340;
Best Local Similarity 26.8%; Pred. No. 9.1e-06;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;

Qy      6 QPVCRGG---TQRPCYKVIYFHDTSRRLNFEAKACRRDGGQVLVSIESDEQKLIKFP 62
      : : : : : : : : : : : : : : : : : : : : : : : :
Db      1127 QKLCEGWTKFGHCYR--HFPD---RATWVDAESQCRKQOQSHLSIVTPEQ---EFV 1177
      : : : : : : : : : : : : : : : : : : : : : : : :

Qy      63 ENLLPSDGF-WIGLRREEKQSNSTACQDLVATDGSISQFRNMYVDEP-----SCGSEV 117
      : : : : : : : : : : : : : : : : : : : : : : : :
Db      1178 NN---NAQDYQWIGL-----NDKTTIEGDFRMSDGHSLQFNWRPNQDNFATGEDC 1256
      : : : : : : : : : : : : : : : : : : : : : : : :

Qy      118 CVVWYHPSAPAGTGGPYMFQWDDRCNNWNNFIKYS-----DEKPAVPSREAGEETE 172
      : : : : : : : : : : : : : : : : : : : : : : : :
Db      1227 VVMTIWEKSG-----EWNDVPCNYQLPFTCKGTGTVACGEPVVEHARIFGKKD 1274
      : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 12
T42630
aggreCAN - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42630
R:Harig, T.M.; Kollar, J.; Huynh, T.D.
submitted to the EMBL Data Library, September 1996
A:Description: Complete coding sequence of bovine aggreCAN: comparative structural an
A:Reference number: Z22182
A:Accession: T42630
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-2327 <HR>
A:Cross-references: EMBL:U76615; NID:G1730259; PID:G1730260; PIDN:AAB38524.1
A:Experimental source: articular chondrocytes
C:Superfamily: aggreCAN; C-type lectin homology; complement factor H repeat homology;
C:Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glycop
Query Match 14.4% Score 158.5; DB 2; Length 2327;
Best Local Similarity 26.8%; Pred. No. 1.7e-05;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;
QY 6 QPVCRRG--TORPCYKVIYFHDTSRLNPFREAKACRRDGGQLVYSIESEDEQKLIETKFI 62
Db 2114 QKICEEGWTFFQGHCHYR--HPFD---RATWDAESQCRKQSHLSIVTPEQ-----EPV 2164
QY 63 ENLLPSDGF-WIGLRREEKQSNSTACQDLYAWTDGSIQFRNYYVDEP-----SCGSEV 117
Db 2165 NN---NAQDYQWIGL-----NDKTIEGDFRWSGDGSLQFNWRPNQDPNFFATGDC 2213
QY 118 CVVMYHQPSAPAGIGGPMFQWDDRCNMKNPFTCKYS-----DEKPAVPSRAEGSETE 172
Db 2214 VVMVIEHGK-----EWNDVPCNYQLPFTCKGTACGEPVVEHARIFGOKKD 2261

```

RESULT 13
A39086
aggrecan precursor, cartilage long splice form [validated] - human
N;Alternate names: chondroitin sulfate proteoglycan 1; large aggregating proteoglycan;
N;Contains: aggrecan cartilage short splice form
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence revision 01-Dec-2000 #text change 08-Dec-2000

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:07:10 ; Search time 19.3982 Seconds
(without alignments)
1954.412 Million cell updates/sec

Title: US-09-887-855-2_COPY_25_227

Perfect score: 1100

Sequence: 1 RLLSGQPVCRGRTORPCYKV.....BEDAKTFKESREALNLAY 203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 1100 | 100.0 | 206 | 9 | US-09-887-855-5 |
| 2 | 1100 | 100.0 | 374 | 9 | US-09-887-855-2 |
| 3 | 1100 | 100.0 | 374 | 12 | US-10-094-749-2090 |
| 4 | 1100 | 100.0 | 374 | 15 | US-10-149-819-15 |
| 5 | 1096 | 99.6 | 374 | 12 | US-10-094-749-2142 |
| 6 | 1086 | 98.7 | 382 | 10 | US-09-909-320-137 |
| 7 | 1086 | 98.7 | 382 | 10 | US-09-909-088-137 |
| 8 | 1086 | 98.7 | 382 | 10 | US-09-905-291A-137 |
| 9 | 1086 | 98.7 | 382 | 10 | US-09-902-853-137 |
| 10 | 1086 | 98.7 | 382 | 10 | US-09-907-824-137 |
| 11 | 1086 | 98.7 | 382 | 10 | US-09-907-841-137 |
| 12 | 1086 | 98.7 | 382 | 11 | US-09-904-011-137 |
| 13 | 1086 | 98.7 | 382 | 11 | US-09-906-742-137 |
| 14 | 1086 | 98.7 | 382 | 11 | US-09-906-838-137 |
| 15 | 1086 | 98.7 | 382 | 11 | US-09-907-613-137 |

| | | | | | | |
|----|------|------|-----|----|--------------------|-------------------|
| 16 | 1086 | 98.7 | 382 | 11 | US-09-907-942-137 | Sequence 137, App |
| 17 | 1086 | 98.7 | 382 | 11 | US-09-904-859-137 | Sequence 137, App |
| 18 | 1086 | 98.7 | 382 | 11 | US-09-909-204-137 | Sequence 137, App |
| 19 | 1086 | 98.7 | 382 | 11 | US-09-904-820-137 | Sequence 137, App |
| 20 | 1086 | 98.7 | 382 | 11 | US-09-904-786-137 | Sequence 137, App |
| 21 | 1086 | 98.7 | 382 | 11 | US-09-906-646-137 | Sequence 137, App |
| 22 | 1086 | 98.7 | 382 | 11 | US-09-906-700-137 | Sequence 137, App |
| 23 | 1086 | 98.7 | 382 | 11 | US-09-903-786-137 | Sequence 137, App |
| 24 | 1086 | 98.7 | 382 | 11 | US-09-902-803-137 | Sequence 137, App |
| 25 | 1086 | 98.7 | 382 | 11 | US-09-903-749A-137 | Sequence 137, App |
| 26 | 1086 | 98.7 | 382 | 11 | US-09-904-119-137 | Sequence 137, App |
| 27 | 1086 | 98.7 | 382 | 11 | US-09-904-956-137 | Sequence 137, App |
| 28 | 1086 | 98.7 | 382 | 11 | US-09-902-736-137 | Sequence 137, App |
| 29 | 1086 | 98.7 | 382 | 11 | US-09-907-794-137 | Sequence 137, App |
| 30 | 1086 | 98.7 | 382 | 11 | US-09-903-943-137 | Sequence 137, App |
| 31 | 1086 | 98.7 | 382 | 11 | US-09-904-462-137 | Sequence 137, App |
| 32 | 1086 | 98.7 | 382 | 11 | US-09-907-925-137 | Sequence 137, App |
| 33 | 1086 | 98.7 | 382 | 11 | US-09-902-692-137 | Sequence 137, App |
| 34 | 1086 | 98.7 | 382 | 11 | US-09-903-520-137 | Sequence 137, App |
| 35 | 1086 | 98.7 | 382 | 11 | US-09-905-056-137 | Sequence 137, App |
| 36 | 1086 | 98.7 | 382 | 11 | US-09-909-064-137 | Sequence 137, App |
| 37 | 1086 | 98.7 | 382 | 11 | US-09-904-553-137 | Sequence 137, App |
| 38 | 1086 | 98.7 | 382 | 11 | US-09-905-381-137 | Sequence 137, App |
| 39 | 1086 | 98.7 | 382 | 11 | US-09-905-088-137 | Sequence 137, App |
| 40 | 1086 | 98.7 | 382 | 11 | US-09-907-575-137 | Sequence 137, App |
| 41 | 1086 | 98.7 | 382 | 11 | US-09-905-075-137 | Sequence 137, App |
| 42 | 1086 | 98.7 | 382 | 11 | US-09-902-759-137 | Sequence 137, App |
| 43 | 1086 | 98.7 | 382 | 11 | US-09-902-634-137 | Sequence 137, App |
| 44 | 1086 | 98.7 | 382 | 11 | US-09-902-713-137 | Sequence 137, App |
| 45 | 1086 | 98.7 | 382 | 11 | US-09-907-979-137 | Sequence 137, App |

ALIGNMENTS

RESULT 1

US-09-887-855-5
; Sequence 5, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-887-855-5

| | | | | |
|-----------------------|-----------------|--|-----------|-------------|
| Query Match | 100.0%; | Score 1100; | DB 9; | Length 206; |
| Best Local Similarity | 100.0%; | Pred. No. 3.3e-103; | | |
| Matches 203; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy | 1 | RLLSGQPVCRGRTORPCYKVYFHDTSRRLNFEBAKEACRRDGGQLVSI | SEDEQKLIK | 60 |
| Db | 4 | RLLSGQPVCRGRTORPCYKVYFHDTSRRLNFEBAKEACRRDGGQLVSI | SEDEQKLIK | 63 |
| Qy | 61 | FIENLLPSDGDGFWIGLRRREKQSNSTACODLYAWTDSISQFRNYYVDEPSCGSEVCW | | 120 |
| Db | 64 | FIENLLPSDGDGFWIGLRRREKQSNSTACODLYAWTDSISQFRNYYVDEPSCGSEVCW | | 123 |
| Qy | 121 | MYHPSAPAGTGGPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPVLPE | | 180 |
| Db | 124 | MYHPSAPAGTGGPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPVLPE | | 183 |
| Qy | 181 | ETQEDAKTKFKESREALNLAY | | 203 |
| Db | 184 | ETQEDAKTKFKESREALNLAY | | 206 |


```

RESULT 2
US-09-887-855-2
; Sequence 2, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-887-855-2

Query Match      100.0%; Score 1100; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 7.3e-103;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy      1  RLLSGQPVCRGGTQPCVKVIYFHDTSRRLNFEFEAKEACRRDGGOLVSISEDEOKLIEK 60
Db      25  RLLSGQPVCRGGTQPCVKVIYFHDTSRRLNFEFEAKEACRRDGGOLVSISEDEOKLIEK 84

Qy      61  FIENLLPSDGDFTWGLRRREKQSNSTACQDLYAWTDGSIISQFRNYYVDSPSCGSEVCW 120
Db      85  FIENLLPSDGDFTWGLRRREKQSNSTACQDLYAWTDGSIISQFRNYYVDSPSCGSEVCW 144

Qy      121 MYHOPSPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPVLPE 180
Db      145 MYHOPSPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPVLPE 204

Qy      181 ETQEDAKKTFKESREALNLAY 203
Db      205 ETQEDAKKTFKESREALNLAY 227

```

RESULT 3
US-10-094-749-2090
; Sequence 2090, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAO RU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14

```

; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2090
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2090

Query Match      100.0%; Score 1100; DB 12; Length 374;
Best Local Similarity 100.0%; Pred. No. 7.3e-103;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      1 RLLSGQPCVCGGTGTCVKYIFHDTSRRLNFEEAKEACRRDGGQLVSISEDEQKLTIEK 60
DB      25 RLLSGQPCVCGGTGTCVKYIFHDTSRRLNFEEAKEACRRDGGQLVSISEDEQKLTIEK 84

QY      61 FIENLLPSDGDFTWGLRRREEKQSNSTACQDLYAWTDGSIQSFRNYYVDPSFCGSEVCW 120
DB      85 FIENLLPSDGDFTWGLRRREEKQSNSTACQDLYAWTDGSIQSFRNYYVDPSFCGSEVCW 144

QY      121 MYHOPSPAGIGGGYMFQWMDRRNCMKNFICKYSDEKPAVPSREAGEETELTTPVLPE 180
DB      145 MYHOPSPAGIGGGYMFQWMDRRNCMKNFICKYSDEKPAVPSREAGEETELTTPVLPE 204

QY      181 ETQEDAKTTFKESREAAALNLAY 203
DB      205 ETQEDAKTTFKESREAAALNLAY 227

RESULT 4
US-10-149-819-15
; Sequence 15, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariyah R.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: SHAH, Purvi
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PP-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 3143411CD1
US-10-149-819-15

```

| | Query Match | 100.0%; | Score 1100; | DB 15; | Length 374; |
|----|-----------------------|---|---------------------|-----------|-------------|
| | Best Local Similarity | 100.0%; | Pred. No. 7.3e-103; | | |
| | Matches 203; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| Qy | 1 | RLLSGQPVCRGGTQRCYKVIYFHDTSRRLNPFEEAKACRRDGGQLVSTIESEDEQKLEK | 60 | | |
| | | | | | |
| Db | 25 | RLLSGQPVCRGGTQRCYKVIYFHDTSRRLNPFEEAKACRRDGGQLVSTIESEDEQKLEK | 84 | | |
| | | | | | |
| Qy | 61 | FIEENLIPSGDFFWIGLRBRREEKQSNSTACODLYAWTDGSIQPRNNYVDEPSCGSEVCVV | 120 | | |
| | | | | | |
| Db | 85 | FIEENLIPSGDFFWIGLRBRREEKQSNSTACODLYAWTDGSIQPRNNYVDEPSCGSEVCVV | 144 | | |
| | | | | | |

Qy 121 MYHPSAPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPVLPE 180
Db 145 MYHPSAPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPVLPE 204
Qy 181 ETQEDAKTKPKESREAAINLAY 203
Db 205 ETQEDAKTKPKESREAAINLAY 227

RESULT 5
US-10-094-749-2142
; Sequence 2142, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2142
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2142

Query Match 99.6%; Score 1096; DB 12; Length 374;
Best Local Similarity 99.5%; Pred. No. 1.8e-102;
Matches 202; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 RLLSGQVRCGGTQPCYKVIYFHDTSRRLLNFEEAKACRRDGGQLVSISEDEQKLIK 60
Db 25 RLLSGQVRCGGTQPCYKVIYFHDTSRRLLNFEEAKACRRDGGQLVSISEDEQKLIK 84
Qy 61 FIENLLPSDGDFTWGLRRREKQSNSTACODLYAWTDGSIQFNNWYVDPSGSEYCVV 120
Db 85 FIENLLPSDGDFTWGLRRREKQSNSTACODLYAWTDGSIQFNNWYVDPSGSEYCVV 144
Qy 121 MYHPSAPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPVLPE 180
Db 145 MYHPSAPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTTPVLPE 204
Qy 181 ETQEDAKTKPKESREAAINLAY 203
Db 205 ETQEDAKTKPKESREAAINLAY 227

RESULT 6
US-09-909-320-137
; Sequence 137, Application US/09909320

; Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Deonoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-320-137

, PRIOR FILING DATE: 1999-10-05
 , PRIOR APPLICATION NUMBER: PCT/US99/28214
 , PRIOR FILING DATE: 1999-11-29
 , PRIOR APPLICATION NUMBER: PCT/US99/28313
 , PRIOR FILING DATE: 1999-11-30
 , PRIOR APPLICATION NUMBER: PCT/US99/28564
 , PRIOR FILING DATE: 1999-12-02
 , PRIOR APPLICATION NUMBER: PCT/US99/28565
 , PRIOR FILING DATE: 1999-12-02
 , PRIOR APPLICATION NUMBER: PCT/US99/30095
 , PRIOR FILING DATE: 1999-12-16
 , PRIOR APPLICATION NUMBER: PCT/US99/30911
 , PRIOR FILING DATE: 1999-12-20
 , PRIOR APPLICATION NUMBER: PCT/US99/30999
 , PRIOR FILING DATE: 1999-12-20
 , PRIOR APPLICATION NUMBER: PCT/US00/00219
 , PRIOR FILING DATE: 2000-01-05
 , NUMBER OF SEQ ID NOS: 423
 , SEQ ID NO 137

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-088B-137

Query Match      98.7%; Score 1086; DB 10; Length 382;
Best Local Similarity 96.2%; Pred. No. 2e-101;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 RLSS-----GQVCRGGTORPCYKVIFYFHDTSRLNPFEEAKEACRRDGGOLVSI 52
    |||||
Db 25 RLASDLRLGGQPVCRGGTORPCYKVIFYFHDTSRLNPFEEAKEACRRDGGOLVSI 84
    |||||

QY 53 DEQKLEKFIENLLPSDGFWGLRRREEKQSNSTACQDLYAWTDGSIQFRNWWYDEPS 112
    |||||
Db 85 DEQKLEKFIENLLPSDGFWGLRRREEKQSNSTACQDLYAWTDGSIQFRNWWYDEPS 144
    |||||

QY 113 CGSEVCVMYHQPSAPAGIGGPPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETE 172
    |||||
Db 145 CGSEVCVMYHQPSAPAGIGGPPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETE 204
    |||||

QY 173 LTTPLVLPETQBEDAKKTFKESREALNLAY 203
    |||||
Db 205 LTTPLVLPETQBEDAKKTFKESREALNLAY 235
    |||||

RESULT 8
US-09-905-291A-137
; Sequence 137, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tysse, Daniel

```

```

; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-905-291A-137

Query Match          98.7%; Score 1086; DB 10; Length 382;
Best Local Similarity 96.2%; Pred. No. 2e-101;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 RLLS-----GQPVCRGCTGTCPCYKVIYFHDTSRLNFEFAKEACRDRGGQLVSIET 52
Db 25 RLLSASDLDRGQPCVCRGCTGTCPCYKVIYFHDTSRLNFEFAKEACRDRGGQLVSIET 84
QY 53 DRQKLEKTIENLLPSDGFHWGLRRREKQSNSTACQDLAWTDCGSIQFRNYYVDEPS 112
Db 85 DRQKLEKTIENLLPSDGFHWGLRRREKQSNSTACQDLAWTDCGSIQFRNYYVDEPS 144
QY 113 CGSEVCVMYHOPSPAGTGGPYMFQWDDRCNMKNFICKYSDKPAVPSRAGEETE 172
Db 145 CGSEVCVMYHOPSPAGTGGPYMFQWDDRCNMKNFICKYSDKPAVPSRAGEETE 204
QY 173 LTPVLPETQEDAKTKPKESREAAALNLAY 203
Db 205 LTPVLPETQEDAKTKPKESREAAALNLAY 235

RESULT 9
US-09-902-853-137
; Sequence 137, Application US/09902853
; Publication No. US20020192659A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Deanoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Goddard, A.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,853
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US/09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-902-853-137

Query Match          98.7%; Score 1086; DB 10; Length 382;
```

```
Best Local Similarity 96.2%; Pred. No. 2e-101; Mismatches 0; Indels 8; Gaps 1;
Matches 203; Conservative 0;

QY 1 RLLS-----GQVCRGGTQPCVKIYFHDTSRRLNFEAKACRRDGGQLVSIIE 52
    |||||
Db 25 RLLSASDLRLRGQVCRGGTQPCVKIYFHDTSRRLNFEAKACRRDGGQLVSIIE 84
    |||||
QY 53 DEQKLEKFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS 112
    |||||
Db 85 DEQKLEKFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS 144
    |||||
QY 113 CGSEVCVMYHQPSPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAGEETE 172
    |||||
Db 145 CGSEVCVMYHQPSPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAGEETE 204
    |||||
QY 173 LTPVLPETQEDAKKTFKESREAAALNLAY 203
    |||||
Db 205 LTPVLPETQEDAKKTFKESREAAALNLAY 235
    |||||

RESULT 10
US-09-907-824-137
; Sequence 137, Application US/09907824
; Publication No. US20020197671A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,824
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
```

```
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-907-824-137

Query Match 98.7%; Score 1086; DB 10; Length 382;
Best Local Similarity 96.2%; Pred. No. 2e-101;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 RLLS-----GQVCRGGTQPCVKIYFHDTSRRLNFEAKACRRDGGQLVSIIE 52
    |||||
Db 25 RLLSASDLRLRGQVCRGGTQPCVKIYFHDTSRRLNFEAKACRRDGGQLVSIIE 84
    |||||
QY 53 DEQKLEKFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS 112
    |||||
Db 85 DEQKLEKFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS 144
    |||||
QY 113 CGSEVCVMYHQPSPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAGEETE 172
    |||||
Db 145 CGSEVCVMYHQPSPAGIGGPMFQWDDRCNMKNFICKYSDEKPAVPSREAGEETE 204
    |||||
QY 173 LTPVLPETQEDAKKTFKESREAAALNLAY 203
    |||||
Db 205 LTPVLPETQEDAKKTFKESREAAALNLAY 235
    |||||

RESULT 11
US-09-907-841-137
; Sequence 137, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
```

```

; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,841
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-907-841-137

Query Match      98.7%; Score 1086; DB 10; Length 382;
Best Local Similarity 96.2%; Pred. No. 2e-101;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY      1 RLLS-----GQPCVCGGTQPCVKVIYFHDTSRLNPFEEAKEACRRDGGQGVSE 52
      |||||
Db      25 RLLSASDLRLGCGQPCVCGGTQPCVKVIYFHDTSRLNPFEEAKEACRRDGGQGVSE 84
      |||||
QY      53 DEQKLEKFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPS 112
      |||||
Db      85 DEQKLEKFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPS 144
      |||||
QY      113 CGSEVCVMYHQSPAPAGIGGPMFQWDDRCNMKNFKICKYSDEKPAVPSREAGEETE 172
      |||||
Db      145 CGSEVCVMYHQSPAPAGIGGPMFQWDDRCNMKNFKICKYSDEKPAVPSREAGEETE 204
      |||||
QY      173 LTPVLPPEETQEDAKKTFKESREAAALNAY 203
      |||||
Db      205 LTPVLPPEETQEDAKKTFKESREAAALNAY 235
      |||||

RESULT 12
US-09-011-137
; Sequence 137, Application US/09904011
; Publication No. US2003000330A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
```

```

; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-904-011-137

Query Match      98.7%; Score 1086; DB 11; Length 382;
Best Local Similarity 96.2%; Pred. No. 2e-101;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY      1 RLLS-----GQPCVCGGTQPCVKVIYFHDTSRLNPFEEAKEACRRDGGQGVSE 52
      |||||
Db      25 RLLSASDLRLGCGQPCVCGGTQPCVKVIYFHDTSRLNPFEEAKEACRRDGGQGVSE 84
      |||||
QY      53 DEQKLEKFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPS 112
      |||||
Db      85 DEQKLEKFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPS 144
      |||||
```

| | | | |
|----|-----|---------------------------------------|-----|
| Qy | 113 | CGSEVCVVVYHOPSPAGITGGPYMFQWDDRCNNKNPI | 172 |
| | | CKYSDKPAVPSREAGEETE | |
| Db | 145 | CGSEVCVVVYHOPSPAGITGGPYMFQWDDRCNNKNPI | 204 |
| | | CKYSDKPAVPSREAGEETE | |
| Qy | 173 | LTTPVLPPEETOEDAKTKFKESREAAALNAY | 203 |
| Db | 205 | LTTPVLPPEETOEDAKTKFKESREAAALNAY | 235 |

RESULT 13

US-09-906-742-137

Sequence 137, Application US/09906742

Publication No. US20030023054A1

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: Ashkenazi, Avi.

APPLICANT: Botstein, David

APPLICANT: Desnoyers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerber, Hanspeter

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, A.

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth, J.

APPLICANT: Kljavin, Ivar J.

APPLICANT: Mather, Jennie P.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William, I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

TITLE OF INVENTION: Acids Encoding the Same

FILE REFERENCE: 10466-14

CURRENT APPLICATION NUMBER: US/09/906,742

CURRENT FILING DATE: 2001-07-16

PRIOR FILING DATE: 2000-09-18

PRIOR APPLICATION NUMBER: PCT/US00/04414

PRIOR FILING DATE: 2000-02-22

PRIOR APPLICATION NUMBER: US 60/143,048

PRIOR FILING DATE: 1999-07-07

PRIOR APPLICATION NUMBER: US 60/145,698

PRIOR FILING DATE: 1999-07-26

PRIOR APPLICATION NUMBER: US 60/146,222

PRIOR FILING DATE: 1999-07-28

PRIOR APPLICATION NUMBER: PCT/US99/20594

PRIOR FILING DATE: 1999-09-08

PRIOR APPLICATION NUMBER: PCT/US99/20944

PRIOR FILING DATE: 1999-09-13

PRIOR APPLICATION NUMBER: PCT/US99/21090

PRIOR FILING DATE: 1999-09-15

PRIOR APPLICATION NUMBER: PCT/US99/21547

PRIOR FILING DATE: 1999-09-15

PRIOR APPLICATION NUMBER: PCT/US99/23089

PRIOR FILING DATE: 1999-10-05

PRIOR APPLICATION NUMBER: PCT/US99/28214

PRIOR FILING DATE: 1999-11-29

PRIOR APPLICATION NUMBER: PCT/US99/28313

PRIOR FILING DATE: 1999-11-30

PRIOR APPLICATION NUMBER: PCT/US99/28564

PRIOR FILING DATE: 1999-12-02

PRIOR APPLICATION NUMBER: PCT/US99/28565

PRIOR FILING DATE: 1999-12-02

```

; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-906-742-137

Query Match          98.7%; Score 1086; DB 11; Length 382;
Best Local Similarity 96.2%; Pred. No. 2e-101;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 RLLS-----GQPCVCGTQRPCYKVIFYFHDTSRLNPFEEAKCRDGGQLVSI 52
   |||||
Db 25 RLLSASDLDLRGQPCVCGTQRPCYKVIFYFHDTSRLNPFEEAKCRDGGQLVSI 84
   |||||
Qy 53 DEQKLTIEKPIENLLPSGDGFWIGLRREEKQSNSTACQDLAYWTDGSI 112
   |||||
Db 85 DEQKLTIEKPIENLLPSGDGFWIGLRREEKQSNSTACQDLAYWTDGSI 144
   |||||
Qy 113 CGSEVCCVMYHQPSAPAGIGGPPYMFQWDDRCNMKNFKICKYDSEKPAVPSREABGEETE 172
   |||||
Db 145 CGSEVCCVMYHQPSAPAGIGGPPYMFQWDDRCNMKNFKICKYDSEKPAVPSREABGEETE 194
   |||||
Qy 173 LTTPVLPEETQEDDAKKTFKESREAAINLAY 203
   |||||
Db 205 LTTPVLPEETQEDDAKKTFKESREAAINLAY 235
   |||||

```

RESULT 14

US-09-906-838-137

Sequence 137, Application US/09906838

Publication No. US20030027143A1

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: Ashkenazi, Avi

APPLICANT: Botstein, David

APPLICANT: Deconvoyers, Luc

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Filvaroff, Ellen

APPLICANT: Fong, Sherman

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerber, Hanspeter

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, A.

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, Christopher J.

APPLICANT: Gurney, Austin L.

APPLICANT: Hillan, Kenneth, J.

APPLICANT: Kijavlin, Ivar J.

APPLICANT: Mather, Jennie P.

APPLICANT: Pan, James

APPLICANT: Paoni, Nicholas F.

APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William, I.

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

TITLE OF INVENTION: Acids Encoding the Same

FILE REFERENCE: 10466-14

CURRENT APPLICATION NUMBER: US/09/906,838

CURRENT FILING DATE: 2001-07-16

PRIOR APPLICATION NUMBER: 09/665,350

PRIOR FILING DATE: 2000-09-18

APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/907.613
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 137
LENGTH: 382
TYPE: PRT
ORGANISM: Homo sapiens
US-09-906-838-137

Query Match 98.7%; Score 1086; DB 11; Length 382;
Best Local Similarity 96.2%; Pred. No. 2e-101;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 1 RLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSTSE 52
DB 25 RLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSTSE 84
QY 53 DEQKLEKFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNRYVDEPS 112
DB 85 DEQKLEKFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQFRNRYVDEPS 144
QY 113 CGSEVCVMYHQSAPAGIGPWFQWDRCKNKNFIKYSDEKPAVPSRAEGSETE 172
DB 145 CGSEVCVMYHQSAPAGIGPWFQWDRCKNKNFIKYSDEKPAVPSRAEGSETE 204
QY 173 LTPVLPPEETOEDAKKTFKESREAALNLAY 203
DB 205 LTPVLPPEETOEDAKKTFKESREAALNLAY 235

RESULT 15
US-09-907-613-137
Sequence 137, Application US/09907613
Publication No. US20030027145A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnovers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone

Query Match 98.7%; Score 1086; DB 11; Length 382;
Best Local Similarity 96.2%; Pred. No. 2e-101;
Matches 203; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 1 RLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSTSE 52
DB 25 RLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSTSE 84

| | | | | | | | | | | | | | | | | | | | | | | |
|----|-----|--------|-------|-------|------|-------|------|------|------|------|------|------|------|----|------|-----|----|----|----|----|-----|-----|
| Qy | 53 | DEQKLI | EKFIE | NLLPS | DCGF | WIGLR | REEK | QSGN | STAC | ODLY | AWTD | GSIS | QFPR | VV | DEDS | 112 | | | | | | |
| Db | 85 | DEQKLI | EKFIE | NLLPS | DCGF | WIGLR | REEK | QSGN | STAC | ODLY | AWTD | GSIS | QFPR | VV | DEDS | 144 | | | | | | |
| Qy | 113 | CGSEV | CVVM | YHQ | PSAP | AGIG | PGPY | MQW | ND | DR | CNM | KNF | ICK | YS | DSKP | VP | SR | EA | GE | ET | 172 | |
| Db | 145 | CGSEV | CVVM | YHQ | PSAP | AGIG | PGPY | MQW | ND | DR | CNM | KNF | ICK | YS | DSKP | VP | SR | EA | GE | ET | 204 | |
| Qy | 173 | LTTPV | LP | PE | TQ | ED | AK | TF | KGS | R | E | A | L | N | L | A | L | N | L | A | L | 203 |
| Db | 205 | LTTPV | LP | PE | TQ | ED | AK | TF | KGS | R | E | A | L | N | L | A | L | N | L | A | L | 235 |

Search completed: December 22, 2003, 16:16:48
Job time : 20.3982 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:05:09 ; Search time 6.4979 Seconds
(without alignments)
1476.391 Million cell updates/sec

Title: US-09-887-855-2_COPY_24_227

Perfect score: 1106

Sequence: 1 GRLLSGQPVCRGRTQPCYK.....BEDAKTKFKESREALNLAY 204

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 574 | 51.9 | 273 | 1 | CHOD MOUSE |
| 2 | 561.5 | 50.8 | 273 | 1 | CHOD_HUMAN |
| 3 | 182 | 16.5 | 1456 | 1 | MANR_HUMAN |
| 4 | 177 | 16.0 | 1268 | 1 | PGCN_MOUSE |
| 5 | 176 | 15.9 | 1321 | 1 | PGCN_HUMAN |
| 6 | 174.5 | 15.8 | 3381 | 1 | PGCV_BOVIN |
| 7 | 174 | 15.7 | 1257 | 1 | PGCV_RAT |
| 8 | 174 | 15.7 | 2738 | 1 | PGCV_RAT |
| 9 | 174 | 15.7 | 3358 | 1 | PGCV_MOUSE |
| 10 | 174 | 15.7 | 3396 | 1 | PGCV_HUMAN |
| 11 | 171 | 15.5 | 3562 | 1 | PGCV_CHICK |
| 12 | 165 | 14.9 | 643 | 1 | CD93_RAT |
| 13 | 158.5 | 14.3 | 2364 | 1 | PGCA_BOVIN |
| 14 | 158.5 | 14.3 | 2415 | 1 | PGCA_HUMAN |
| 15 | 155.5 | 14.1 | 2333 | 1 | PGCA_CANFA |
| 16 | 154.5 | 14.0 | 612 | 1 | LEM2_MOUSE |
| 17 | 153.5 | 13.9 | 644 | 1 | CD93_MOUSE |
| 18 | 153.5 | 13.9 | 2124 | 1 | PGCA_RAT |
| 19 | 152 | 13.7 | 912 | 1 | PCGB_BOVIN |
| 20 | 151.5 | 13.7 | 652 | 1 | CD93_HUMAN |
| 21 | 151 | 13.7 | 197 | 1 | CUF1_HUMAN |
| 22 | 149 | 13.5 | 321 | 1 | FCB2_HUMAN |
| 23 | 147.5 | 13.3 | 2132 | 1 | PGCA_MOUSE |
| 24 | 147 | 13.3 | 883 | 1 | PCGB_MOUSE |
| 25 | 146 | 13.2 | 2109 | 1 | PGCA_CHICK |
| 26 | 145 | 13.1 | 158 | 1 | LECG_TRIST |
| 27 | 145 | 13.1 | 883 | 1 | PCGB_RAT |
| 28 | 144.5 | 13.1 | 173 | 1 | LEC2_MEGRO |
| 29 | 144.5 | 13.1 | 372 | 1 | LEM1_RAT |
| 30 | 143.5 | 13.0 | 372 | 1 | LEM1_MOUSE |
| 31 | 141.5 | 12.8 | 331 | 1 | FCE2_MOUSE |
| 32 | 141.5 | 12.8 | 549 | 1 | LEM2_RAT |
| 33 | 140.5 | 12.7 | 162 | 1 | LEC3_MEGRO |

| | | | | | |
|----|-------|------|-----|---|------------|
| 34 | 140 | 12.7 | 370 | 1 | LEM1_BOVIN |
| 35 | 139 | 12.6 | 248 | 1 | PSPA_HUMAN |
| 36 | 139 | 12.6 | 283 | 1 | LECA_SARPE |
| 37 | 138.5 | 12.5 | 152 | 1 | IXA_TRIFL |
| 38 | 138.5 | 12.5 | 202 | 1 | TETN_MOUSE |
| 39 | 137.5 | 12.4 | 372 | 1 | LEM1_MACMU |
| 40 | 137.5 | 12.4 | 372 | 1 | LEM1_PAPHA |
| 41 | 136 | 12.3 | 175 | 1 | LITH_BOVIN |
| 42 | 135 | 12.2 | 165 | 1 | LITI_MOUSE |
| 43 | 134.5 | 12.2 | 132 | 1 | ACAL_ANSAN |
| 44 | 134.5 | 12.2 | 372 | 1 | LEM1_PONPY |
| 45 | 134.5 | 12.2 | 485 | 1 | LEM2_BOVIN |

ALIGNMENTS

RESULT 1

| ID | CHOD_MOUSE | STANDARD; | PRT; | 273 AA. |
|----|--|-----------|------|---------|
| AC | O9CXM0: Q8V131; | | | |
| DT | 28-FEB-2003 (Rel. 41, Created) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last sequence update) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | | |
| DE | Chondrolectin precursor (Transmembrane protein MT75). | | | |
| GN | CHODL | | | |
| OS | Mus musculus (Mouse) | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | |
| OX | NCBI_TaxID=10090; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=C57BL/6J; | | | |
| RA | Weng L., Smith P., Hubner R., Wouters J., Merregaert J.; | | | |
| RT | "Mt75, a low expressed c-type lectin gene involving in | | | |
| RT | chondrogenesis." | | | |
| RL | Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases. | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=C57BL/6J; TISSUE=Embryonic head; | | | |
| RX | MEDLINE=21085660; PubMed=11217851; | | | |
| RA | Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., | | | |
| RA | Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., | | | |
| RA | Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., | | | |
| RA | Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., | | | |
| RA | Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., | | | |
| RA | Fleischmann W., Gaasterland T., Giesi C., King B., Kochiwa H., | | | |
| RA | Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., | | | |
| RA | Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., | | | |
| RA | Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., | | | |
| RA | Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., | | | |
| RA | Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., | | | |
| RA | Guatinchich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., | | | |
| RA | Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., | | | |
| RA | Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., | | | |
| RA | Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., | | | |
| RA | Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., | | | |
| RA | Yuzhakov-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., | | | |
| RA | Hayashizaki Y.; | | | |
| RT | "Functional annotation of a full-length mouse cDNA collection."; | | | |
| RL | Nature 409:685-690(2001). | | | |
| CC | -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential). | | | |
| CC | -!- SIMILARITY: Contains 1 C-type lectin family domain. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; AF311699; AAL50354.1; -. | | | |

DR EMBL; AK014255; BAB29226.1; --
 DR HSSP; P22897; IEGG
 DR MGD; MGI:2179069; Chodl.
 DR InterPro; IPR001304; Lectin_C.
 DR Pfam; PF00059; lectin_c; 1.
 DR SMART; SM00034; CLECT; 1.
 DR DR PROSITE; PS00615; C-TYPE LECTIN 1; FALSE_NEG.
 DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.
 KW Lectin; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 273 CHONDROLECTIN.
 FT DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 217 237 POTENTIAL.
 FT DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 35 179 C-TYPE LECTIN
 FT CARBOHYD 86 86 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 24 24 V -> W (IN REF. 2).
 FT CONFLICT 179 179 T -> K (IN REF. 2).
 SQ SEQUENCE 273 AA; 30303 MW; E052D933F244F4C7 CRC64;
 Query Match 51.9%; Score 574; DB 1; Length 273;
 Best Local Similarity 57.4%; Pred. No. 4e-44;
 Matches 109; Conservative 26; Mismatches 41; Indels 14; Gaps 5;
 QY 2 RLISGQPVCRGRTQPCYKVIYHDTSRRLNFEEAKEACRRDGGQGLVSIIESEDEQKLEK 61
 DB 23 RVVSGQKVCADVKHPCYKMYFHELSRVSFQEARLACESEGGVLLSLENEAQKLES 82
 QY 62 FIENLPL-----SDGDFWIGLRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPSCGS 116
 DB 83 MLQNLKPGTGISDGFWIGLRSGDQGT-SGACPDLYQNSDGSQSSQFRNYYVDEPSCGS 141
 QY 117 EVCVMVHQPAPAGIGGPFYMFQWDDRCNMKNFICKYSDE-KPAPVPSREAGEETELT 175
 DB 142 EKCVMVHQTAPNGPLGGLYQNDRCNMKNYICTYPEIHTPEA-----EXPILT 196
 QY 176 TPVLPEETQE 185
 DB 197 NQ--PEETHE 204
 RESULT 2
 CHOD_HUMAN
 ID CHOD_HUMAN STANDARD; PRT; 273 AA.
 AC Q9H9P2; Q9HCY3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 13-SEP-2003 (Rel. 42, Last annotation update)
 DE Chondrolectin precursor (Transmembrane protein MT75) (PRED12 protein).
 DE CHODL OR C21ORF68.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RX MEDLINE=22074930; PubMed=12079284;
 RA Weng L., Smits P., Wauters J., Merregaert J.;
 RT "Molecular cloning and characterization of human chondrolectin, a novel type I transmembrane protein homologous to C-type lectins."; Genomics 80:62-70(2002).
 [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y., Soeda E., Onki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Sakaki T., Nagamine K., Mitsuyama S., Antonarakis S.E., Minoshima S., Shimizu N., Nordtsiek G., Hornischer K., Brandt P.,

RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H., Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21."; Nature 405:311-319(2000).
 [3]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [4]
 RN SEQUENCE OF 27-273 FROM N.A.
 RP Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Wagatsuna M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Ninomiya K., Iwananagi T.;
 RT "NEDO human cDNA sequencing project."; Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 RN TISSUE SPECIFICITY.
 RX MEDLINE=21564202; PubMed=11707072;
 RA Raymond A., Friedli M., Neergard Henriksen C., Chapot F., Deutsch S., Ucia C., Rossier C., Lyle R., Guipponi M., Antonarakis S.E.;
 RT "From PREDS and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map."; Genomics 78:46-54(2001).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- TISSUE SPECIFICITY: Found in spleen, testis, prostate and fetal liver. Expression limited to vascular muscle of testis, smooth muscle of prostate stroma, heart muscle, skeletal muscle, crypts of small intestine, and red pulp of spleen.
 CC -1- PTM: N-glycosylated.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.iesb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 EMBL; AF257472; AAL05981.1; --
 EMBL; AL163217; CAB90388.1; --
 EMBL; BC009418; AB09418.1; --
 EMBL; AK022689; BAB14181.1; ALT_INIT.
 HSSP; P22897; IEGG.
 Genew; HGNC:17807; CHODL.
 MIM; 607247; --
 InterPro; IPR001304; Lectin_C.

DR Pfam; PF00059; lectin_c; 1.
 DR SMART; SM00034; CLECT; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; FALSE_NEG.
 DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.
 KW Lactin; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 273 CHONDROLECTIN.
 FT DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 217 237 POTENTIAL.
 FT DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 35 179 C-TYPE LECTIN.
 FT CARBOHYD 86 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 273 AA; 30431 MW; F48900A0AF572A311 CRC64;

Query Match 50.8%; Score 561.5; DB 1; Length 273;
 Best Local Similarity 60.1%; Pred. No. 5.3e-43;
 Matches 101; Conservative 25; Mismatches 35; Indels 7; Gaps 3;

QY 2 RLASGQVRCGGTQPCYKVIYHDTSRRLNFEAKACRRDGGQLVSISEDEQKLEK 61
 Db 23 RVVSGGKCFADPFRKCYKWAYFHEJSSRVSFQEARLACESEGGVLLSLENAEQKLIES 82

QY 62 FIENLLP-----SDGDFWIGLRREKQSNSTACQDLYAMTDSISQFRNMYVDEPSCGS 116
 Db 83 MLQNLTKPGTGISDGDFFWIGLRNGDQOT-SGACPDLYQWSDGNSQYRNWYTDPESCGS 141

QY 117 EVCVMYHQSAPAGIGGYPYFQWDDRCNMKNPFICKYSDE-KPAVP 163
 Db 142 EKCVMYHPTANPLGLGPPYLYQWDDRCNMKNYICKYPEINPTAP 189

RESULT 3

MANR_HUMAN STANDARD; PRT; 1456 AA.

AC P22897;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Macrophage mannose receptor precursor (MMR) (CD206 antigen).
 GN MRC1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Placenta;
 RA MEDLINE=90324192; PubMed=2373685;
 RA Taylor M.E., Conarty J.T., Lennartz M.R., Stahl P.D., Drickamer K.;
 RT "Primary structure of the mannose receptor contains multiple motifs
 resembling carbohydrate-recognition domains.";
 RL J. Biol. Chem. 265:12156-12162(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93052405; PubMed=1294118;
 RA Kim S.J., Ruiz N., Bezouska K., Drickamer K.;
 RT "Organization of the gene encoding the human macrophage mannose
 receptor (MRC1).";
 RL Genomics 14:721-727(1992).
 RN [3]
 RP STUDIES ON THE BINDING OF INDIVIDUAL LECTIN DOMAINS.
 RX MEDLINE=92112893; PubMed=1730714;
 RA Taylor M.E., Bezouska K., Drickamer K.;
 RT "Contribution to ligand binding by multiple carbohydrate-recognition
 domains in the macrophage mannose receptor.";
 RL J. Biol. Chem. 267:1719-1726(1992).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 642-788.
 RX MEDLINE=20347275; PubMed=10779515;
 RA Feinberg H., Park-Snyder S., Kolatkar A.R., Heise C.T., Taylor M.E.,
 RA Weis W.I.;
 RT "Structure of a C-type carbohydrate recognition domain from the
 macrophage mannose receptor.";

J. Biol. Chem. 275:21539-21548(2000).
 -1- FUNCTION: MEDIATES THE ENDOCYTOSIS OF GLYCOPROTEINS BY
 MACROPHAGES. IN SEVERAL RECOGNITION AND UPTAKE PROCESSES.
 -1- SUBCELLULAR LOCATION: Type I membrane protein.
 -1- MISCELLANEOUS: CRDS 1-3 HAVE AT MOST VERY WEAK AFFINITY FOR
 CARBOHYDRATE. CRD 4 SHOWS THE HIGHEST AFFINITY BINDING AND HAS
 MULTISPECIFICITY FOR A VARIETY OF MONOSACCHARIDES. AT LEAST 3 CRDS
 (4, 5, AND 7) ARE REQUIRED FOR HIGH AFFINITY BINDING AND
 ENDOCYTOSIS OF MULTIVALENT GLYCOCONJUGATES.
 -1- SIMILARITY: Contains 8 C-type lectin family domains.
 -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
 -1- DATABASE: NAME=PROW; NOTE=PROW 2:85-89(2001);
 WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1644341535.g.htm".

 This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).

 EMBL; J05550; AAA59868.1; --
 EMBL; M93221; AAA60389.1; --
 EMBL; M93192; AAA60389.1; JOINED.
 EMBL; M93193; AAA60389.1; JOINED.
 EMBL; M93194; AAA60389.1; JOINED.
 EMBL; M93195; AAA60389.1; JOINED.
 EMBL; M93196; AAA60389.1; JOINED.
 EMBL; M93197; AAA60389.1; JOINED.
 EMBL; M93198; AAA60389.1; JOINED.
 EMBL; M93199; AAA60389.1; JOINED.
 EMBL; M93200; AAA60389.1; JOINED.
 EMBL; M93201; AAA60389.1; JOINED.
 EMBL; M93202; AAA60389.1; JOINED.
 EMBL; M93203; AAA60389.1; JOINED.
 EMBL; M93204; AAA60389.1; JOINED.
 EMBL; M93205; AAA60389.1; JOINED.
 EMBL; M93206; AAA60389.1; JOINED.
 EMBL; M93207; AAA60389.1; JOINED.
 EMBL; M93208; AAA60389.1; JOINED.
 EMBL; M93209; AAA60389.1; JOINED.
 EMBL; M93210; AAA60389.1; JOINED.
 EMBL; M93211; AAA60389.1; JOINED.
 EMBL; M93212; AAA60389.1; JOINED.
 EMBL; M93213; AAA60389.1; JOINED.
 EMBL; M93214; AAA60389.1; JOINED.
 EMBL; M93215; AAA60389.1; JOINED.
 EMBL; M93216; AAA60389.1; JOINED.
 EMBL; M93217; AAA60389.1; JOINED.
 EMBL; M93218; AAA60389.1; JOINED.
 EMBL; M93219; AAA60389.1; JOINED.
 EMBL; M93220; AAA60389.1; JOINED.
 FIR; A36563; A36563.
 PDB; 1EGG; 30-AUG-00.
 PDB; 1EGI; 30-AUG-00.
 Genew; HGNC:7228; MRC1.
 MIM; 153618; --
 GO; GO:0005887; C:integral to plasma membrane; TAS.
 GO; GO:0005537; F:mannose binding activity; TAS.
 GO; GO:0004872; F:receptor activity; TAS.
 GO; GO:0006898; P:receptor mediated endocytosis; TAS.
 InterPro; IPR002353; AntifreezeII.
 InterPro; IPR000562; FN_TypeII.
 InterPro; IPR001304; Lectin_C.
 InterPro; IPR000772; Ricin_B_lectin.
 Pfam; PF00040; fn2; 1.
 Pfam; PF00059; lectin_c; 8.
 Pfam; PF00652; Ricin_B_lectin; 2.
 PRINTS; PR00013; FNTPFELI.
 PRINTS; PR00356; ANTIFREEZEII.
 Prodom; PD000995; FN_Type_II; 1.
 SMART; SM00034; CLECT; 8.

DR SMART; SM00059; FN2; 1.
 DR SMART; SM00458; RICIN; 1.
 DR PROSITE; PS00815; C-TYPE LECTIN 1; 6.
 DR PROSITE; PS00411; C-TYPE LECTIN 2; 8.
 DR PROSITE; PS00023; FIBRONECTIN 2; 1.
 DR PROSITE; PS02311; RICIN B LECTIN; 1.
 DR RECEPTOR; Signal; Calcium-binding; Transmembrane; Repeat;
 KW Glycoprotein; Endocytosis; Macrophage; Lectin; Antigen; 3D-structure.
 FT SIGNAL 1 18
 FT CHAIN 19 1456
 FT DOMAIN 19 1383
 FT TRANSMEM 1384 1411
 FT DOMAIN 1412 1456
 FT DOMAIN 22 142
 FT DOMAIN 157 212
 FT DOMAIN 216 344
 FT DOMAIN 360 490
 FT DOMAIN 502 629
 FT DOMAIN 644 781
 FT DOMAIN 805 926
 FT DOMAIN 943 1083
 FT DOMAIN 1100 1216
 FT DOMAIN 1228 1359
 FT DISULFID 646 659
 FT DISULFID 680 777
 FT DISULFID 753 769
 FT CARBOHYD 104 104
 FT CARBOHYD 344 344
 FT CARBOHYD 529 529
 FT CARBOHYD 926 926
 FT CARBOHYD 930 930
 FT CARBOHYD 1160 1160
 FT CARBOHYD 1205 1205
 FT CARBOHYD 1311 1311
 FT TURN 648 649
 FT STRAND 651 652
 FT TURN 654 655
 FT STRAND 658 663
 FT HELIX 667 669
 FT STRAND 671 671
 FT HELIX 673 683
 FT TURN 684 684
 FT STRAND 686 687
 FT HELIX 693 705
 FT TURN 706 707
 FT TURN 709 710
 FT STRAND 712 718
 FT TURN 723 724
 FT STRAND 727 727
 FT TURN 729 730
 FT STRAND 733 733
 FT TURN 741 742
 FT HELIX 746 748
 FT STRAND 752 757
 FT TURN 758 761
 FT STRAND 764 768
 FT TURN 769 770
 FT STRAND 773 780
 FT TURN 781 782
 SQ SEQUENCE 1456 AA; 166011 MW; 2645EAF3C576A5E3 CRC64;

Query Match 16.5%; Score 182; DB 1; Length 1456;
 Best Local Similarity 25.5%; Pred. NO. 3.3e-08;
 Matches 52; Conservative 37; Mismatches 61; Indels 54; Gaps 8;
 QY 19 YKVIYFHDTSRLNFEAKEACRDGGLVSIQSEDEQKLEKFIENLLPSDGFWIGLR 78
 DB 807 YKYQYFSEKKTMDNARAFCKNFGDLVSIQSESEKFLWKYV-NRNDQAQSYFGLL 865
 QY 79 RREKQSNSTACQDLAYWTDGSIQFRNRYVYDEPSCGS--EVCVVMYHQPSAPAGIGCPY 136
 DB 866 ISLDKK-----FAWMDGSKVDYVSWATGEFNANEDENCVTMY-----SNSGF---- 908

QY 137 MFQWDDRCNMKNPFICKYSDEK---PAVPSREABGEETELTTPVLPEETOE----- 185
 DB 909 ---WNDINCYPNAPICORHSSINATTMP-----TMPSPVSGCKEGNPFYSN 954
 QY 186 -----EDAKTKFKESREAAAL 200
 DB 955 KCPKIFGPMEEERKNWQEARAKACI 978
 RESULT 4
 PGCN_MOUSE
 ID_PGCN_MOUSE STANDARD; PRT; 1268 AA.
 AC P55066;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
 GN CSPG3 OR NCAN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Brain;
 RX MEDLINE=96039250; PubMed=7490074;
 RA Rauch U., Gräpe B., Kulbe G., Arnold-Ammer I., Beier D.,
 RA Faessler R.;
 FT Structure and chromosomal localization of the mouse neurocan gene.";
 RL Genomics 28:405-410(1995).
 CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
 CC development by binding to neural cell adhesion molecules (NG-CAM
 CC and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
 CC acid.
 CC -!- TISSUE SPECIFICITY: BRAIN.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; X84727; CAA59216.1; -.
 CC PIR; S52781; S52781.
 CC HSP; P00740; LEDM.
 CC MGD; MG1:104694; Cspg3.
 CC InterPro; IPR002353; AntifreezeII.
 CC InterPro; IPR000152; Asx_hydroxyl.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF-like.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR003599; Ig.
 CC InterPro; IPR003006; Ig_MHC.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR000538; Link.
 CC InterPro; IPR000436; Sushi_SCR_CCP.
 CC Pfam; PF00008; EGF; 2.
 CC Pfam; PF00047; Ig; 1.
 CC Pfam; PF00059; lectin_c; 1.
 CC Pfam; PF00084; sushi; 1.
 CC Pfam; PF00193; Xlink; 2.
 CC PRINTS; PR01265; LINKMODULE.
 CC PRINTS; PR00356; ANTIFREEZEII.
 CC ProDom; PD000918; Link; 2.

DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX HYDROXYL; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
 DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS00835; IG LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
 KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 1268
 FT DOMAIN 37 157
 FT DOMAIN 158 253
 FT DOMAIN 259 355
 FT DOMAIN 960 996
 FT DOMAIN 998 1034
 FT DOMAIN 1036 1165
 FT DOMAIN 1166 1224
 FT DISULFID 58 139
 FT DISULFID 181 252
 FT DISULFID 205 226
 FT DISULFID 279 354
 FT DISULFID 303 324
 FT DISULFID 964 975
 FT DISULFID 969 984
 FT DISULFID 986 995
 FT DISULFID 1040 1051
 FT DISULFID 1068 1160
 FT DISULFID 1136 1152
 FT DISULFID 1167 1210
 FT DISULFID 1196 1223
 FT CARBOHYD 121 121
 FT CARBOHYD 339 339
 FT CARBOHYD 742 742
 FT CARBOHYD 978 978
 FT CARBOHYD 1175 1175
 SQ SEQUENCE 1268 AA; 137200 MW; 3014E8E202A2FAEC CRC64;

Query Match 16.0%; Score 177; DB 1; Length 1268;
 Best Local Similarity 31.5%; Pred. No. 7.9e-08;
 Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;

Qy 15 QRPCKVLYFHTSRRLNFEAKACRDGGQLVSISEDEKLIKEFIENLLPSDGF 74
 Db 1048 QGHYR--YF---AHRRAWEADRCRRRAGHLTSVHSPEHKFINGF-----GHENSW 1096
 Qy 75 IGLRRREEKQSNSTACQDLVATWDTGSGISQFRNWTYVDPS---CGSEVCVMYHQPSAPAG 131
 Db 1097 IGLNDRTVRD-----FQWTDNTGLQYENWEKQDPNFFAGGDCVVMVAHESG--- 1145
 Qy 132 IGGPYMFQWDRDCRMKNFICK 154
 Db 1146 -----RWNDVPCNLYPVCK 1161

RESULT 5

PGCN HUMAN
 ID PGCN_HUMAN STANDARD; PRT; 1321 AA.
 AC O14594; Q9UPK6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
 GN CSFG3 OR NCAN OR NEUR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99013874; PubMed=9795216;
 RA Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Lennon G.G.;
 RT "Characterization of the human neurocan gene, CSPG3.";
 RL Gene 221:199-205(1998).
 RN [2]
 RP SEQUENCE OF 1-990 AND 1007-1321 FROM N.A.
 RA Lamerdin J.E., McCreedy P.M., Skouronski E., Adamson A.W.,
 RA Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
 RA Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
 RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
 RA Liu S., Attix C., Andreise T., Frankheim M., Amico-Keller G.,
 RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
 RA Krommiller B., Arellano A., Montgomey M., Ow D., Nolan M., Trong S.,
 RA Kobayashi A., Olsen A.S., Carrano A.V.;
 RA "Sequence analysis of an ~1 Mb region containing the MEF2B gene in
 RT 19p12.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
 CC development by binding to neural cell adhesion molecules (NG-CAM
 CC and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
 CC acid.
 CC -!- TISSUE SPECIFICITY: Brain.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF026547; AAC805976.1; -
 DR EMBL; AC003110; AAB86655.1; -
 DR EMBL; AC005254; AAC25581.1; -
 DR HSSP; P00740; 1EDM.
 DR Genew; HGNC:2465; CSPG3.
 DR MIW; 600826; -
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR007110; Ig_Like.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.
 DR PRINTS; PR01265; LINKMODULE.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX HYDROXYL; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
 DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS01187; EGF CA; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1321
FT DOMAIN 38 153
FT DOMAIN 159 254
FT DOMAIN 260 356
FT DOMAIN 1008 1044
FT DOMAIN 1046 1082
FT DOMAIN 1084 1213
FT DOMAIN 1214 1272
FT DISULFID 59 140
FT DISULFID 182 253
FT DISULFID 206 227
FT DISULFID 280 355
FT DISULFID 304 325
FT DISULFID 1012 1023
FT DISULFID 1017 1032
FT DISULFID 1034 1043
FT DISULFID 1088 1099
FT DISULFID 1116 1208
FT DISULFID 1184 1200
FT DISULFID 1215 1258
FT DISULFID 1244 1271
FT CARBOHYD 122 122
FT CARBOHYD 340 340
FT CARBOHYD 1026 1026
FT CARBOHYD 1223 1223
FT CONFLICT 1234 1234
FT CONFLICT 1254 1254
FT CONFLICT 1282 1282
SQ SEQUENCE 1321 AA; 142972 MW; 2EF47F823DB980B8 CRC64;

Query Match
Best Local Similarity 15.9%; Score 176; DB 1; Length 1321;
Matches 45; Conservative 18; Mismatches 48; Indels 32; Gaps 6;

QY 15 QRCYKVIYPHDTSRLNFEAKACRRDGGQLVSEDEOKLIEFKFENLLPSGDGFW 74
DB 1096 QHCYR--YF--AHRAWEDAECRRRSGLHTSVHSPEHSFINSF-----GHENTW 1144

QY 75 IGLRRREKOSNSTACDLYAWTDGSGISQFRNMYVDPS---CGSEVCVVMYHQPSAPAG 131
DB 1145 IGLNDRIVERD-----FQWTDNTGLQFNWRENQPDNFFAGGDCVWVAHESG--- 1193

QY 132 IGGPYFQWNDRCNMKNPFCK 154
DB 1194 -----RWNDVFCVNLPLYVCK 1209

RESULT 6
PGCV BOVIN STANDARD; PRT; 3381 AA.
ID PGCV BOVIN STANDARD; PRT; 3381 AA.
AC P81282; O77610; O77611; O77612;
DT 15-DSC-1998 (Rel. 37, Created)
DT 15-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP).
GN CPBG2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS VO; V1; V2 AND V3).
RC TISSUE=Forebrain;
RC MEDLINE=98288320; PubMed=9624174;

RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.,
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
RL J. Biol. Chem. 273:15758-15764 (1998).
RN [2]
RP SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331
RP AND 342-348
RC TISSUE=Spinal cord;
RX MEDLINE=92062692; PubMed=1720020;
RA Perides G., Biviano F., Bignami A.;
RT "Interaction of a brain extracellular matrix protein with hyaluronic
RT acid.";
RL Biochim. Biophys. Acta 1075:248-258 (1991).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBUNIT: Interacts with PBLN1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=P81282-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=P81282-2; Sequence=VSP_003078, VSP_003079;
CC Name=V2;
CC IsoId=P81282-3; Sequence=VSP_003080;
CC Name=V3;
CC IsoId=P81282-4; Sequence=VSP_003078, VSP_003081;
CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed
CC in the central nervous system, and in a number of mesenchymal and
CC epithelial tissues; the major isoform V2 is restricted to the
CC central nervous system.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (By similarity).
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF060456; AAC24358.1; -;
CC DR EMBL; AF060457; AAC24359.1; -;
CC DR EMBL; AF060458; AAC24360.1; -;
CC DR EMBL; AF060459; AAC24361.1; -;
CC DR F01; T14274; T14274.
CC DR F01; T42389; T42389.
CC DR HSSP; P01132; 1EPG.
CC DR InterPro; IPR000152; Asx_hydroxyl.
CC DR InterPro; IPR000742; EGF-2.
CC DR InterPro; IPR001881; EGF-Ca.
CC DR InterPro; IPR003006; Ig_MHC.
CC DR InterPro; IPR001304; Lectin_C.
CC DR InterPro; IPR000538; Link.
CC DR InterPro; IPR000436; Sushi_SCR_CCP.
CC DR Pfam; PF000008; EGF; 2.
CC DR Pfam; PF00047; Ig; 1.
CC DR Pfam; PF00059; lectin_c; 1.
CC DR Pfam; PF00084; sushi; 1.

DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR PRODOM; PR00356; ANTIFREEZEII.
DR SMART; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 3381
FT DOMAIN 21 147
FT DOMAIN 168 245
FT DOMAIN 266 347
FT DOMAIN 349 1336
FT DOMAIN 1337 3074
FT DOMAIN 3074 3110
FT DOMAIN 3112 3148
FT DOMAIN 3161 3275
FT DOMAIN 3280 3338
FT DISULFID 44 131
FT DISULFID 173 244
FT DISULFID 197 218
FT DISULFID 271 346
FT DISULFID 295 316
FT DISULFID 3078 3089
FT DISULFID 3083 3098
FT DISULFID 3100 3109
FT DISULFID 3116 3127
FT DISULFID 3121 3136
FT DISULFID 3138 3147
FT DISULFID 3154 3165
FT DISULFID 3182 3274
FT DISULFID 3250 3266
FT DISULFID 3281 3324
FT DISULFID 3310 3337
FT CARBOHYD 57 57
FT CARBOHYD 331 331
FT CARBOHYD 352 352
FT CARBOHYD 817 817
FT CARBOHYD 965 965
FT CARBOHYD 1017 1017
FT CARBOHYD 1333 1333
FT CARBOHYD 1393 1393
FT CARBOHYD 1437 1437
FT CARBOHYD 1463 1463
FT CARBOHYD 1653 1653
FT CARBOHYD 1974 1974
FT CARBOHYD 2045 2045
FT CARBOHYD 2074 2074
FT CARBOHYD 2103 2103
FT CARBOHYD 2263 2263
FT CARBOHYD 2290 2290
FT CARBOHYD 2356 2356
FT CARBOHYD 2623 2623
FT CARBOHYD 2641 2641
FT CARBOHYD 2919 2919
FT CARBOHYD 3052 3052
FT CARBOHYD 3354 3354
FT CARBOHYD 3364 3364
FT VARSPLIC 349 349

FT VARSPLIC 350 1336 /FTid=VSP_003078.
FT VARSPLIC 350 1336 Missing (in isoform V1).
FT VARSPLIC 350 1336 /FTid=VSP_003079.
FT VARSPLIC 350 1336 Missing (in isoform V2).
FT VARSPLIC 350 1336 /FTid=VSP_003080.
FT VARSPLIC 350 1336 Missing (in isoform V3).
FT VARSPLIC 350 1336 /FTid=VSP_003081.
FT CONFLICT 25 25 MISSING (IN REF. 2).
FT CONFLICT 51 51 MISSING (IN REF. 2).
FT CONFLICT 89 89 N -> D (IN REF. 2).
FT CONFLICT 96 96 Q -> D (IN REF. 2).
FT CONFLICT 346 346 C -> R (IN REF. 2).
SQ SEQUENCE 3381 AA; 369984 MW; F09716FA7778D459 CRC64;
Query Match 15.8%; Score 174.5; DB 1; Length 3381;
Best Local Similarity 25.5%; Pred. No. 4.4e-07;
Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;
Qy 15 QRPCKYVIYHDTSRRLNFEAKERRDGGQLVSISEDEQKLIKFIEIENLLPSDGF- 73
Db 3162 QGQCYK--YF---AHRRTWDAARECRLOGAHLTSILSHBEQMFVNRV-----GHYQ 3209
Qy 74 WIGLRREEKOSNSTACQDLYAWTSGISQPRNMYVDEP-----SCGSEVCVMYHQSAP 129
Db 3210 WIGL-----NDKMFEDFRWTDGSTLQYENRPNQDSFFSTGDCVVIWHENG-- 3259
Qy 130 AGIGGPFYFOWNDRCKNNFKICKYS-----DEKPAVPSRAEGS----- 170
Db 3260 -----QNDVPCNHLTYTCKGTACQGPVVENAKTFGKMPKRYEINSILRYHC 3310
Qy 171 -----ETELTT-----PVL-----PEETQEDAKTTFKESREAA 201
Db 3311 KDGFIQRHLPTIRCLGNRWAMPKITCLNPSAYQRTYSKKYFKNSSAKDN 3361
RESULT 7
ID PGCN RAT STANDARD; PRT; 1257 AA.
AC P55067;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3)
DE (245 kDa early postnatal core glycoprotein) [Contains: 150 kDa adult
GN CSFG3 OR NCAN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=92406907; PubMed=1326557;
RA Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.;
RT "Cloning and primary structure of neurocan, a developmentally
regulated, aggregating chondroitin sulfate proteoglycan of brain.";
RL J. Biol. Chem. 267:19336-19547(1992).
[2]
CHARACTERIZATION.
RP MEDLINE=94230574; PubMed=7513709;
RX Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.,
RA Margolis R.U., Grumet M.;
RT "The neuronal chondroitin sulfate proteoglycan neurocan binds to the
neural cell adhesion molecules Ng-CAM/Li/NILE and N-CAM, and inhibits
neural adhesion and neurite outgrowth.";
RL J. Cell Biol. 125:669-680(1994).
CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
development by binding to neural cell adhesion molecules (NG-CAM
and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
acid.
CC -!- TISSUE SPECIFICITY: EARLY POSTNATAL AND ADULT BRAIN; NOT EXPRESSED
IN KIDNEY, LUNG, LIVER AND MUSCLE.

CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -1- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
CC DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL
CC BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC -1- SIMILARITY: Contains 2 link domains.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M97161; AAC37679.1; --
CC PIR: S28764; S28764.
CC HSP: P00740; IEDM.
CC DR InterPro: IPR002353; Antifreeze1.
CC DR InterPro: IPR000152; Asx hydroxyl.
CC DR InterPro: IPR000742; EGF_2.
CC DR InterPro: IPR001881; EGF_Ca.
CC DR InterPro: IPR006209; EGF_like.
CC DR InterPro: IPR007110; Ig-like.
CC DR InterPro: IPR003599; Ig.
CC DR InterPro: IPR003006; Ig_MHC.
CC DR InterPro: IPR001304; Lectin_C.
CC DR InterPro: IPR000538; Link.
CC DR InterPro: IPR000436; Sushi_SCR_CCP.
CC DR Pfam: PF00008; EGF; 2.
CC DR Pfam: PF00047; Ig; 1.
CC DR Pfam: PF00059; lectin_c; 1.
CC DR Pfam: PF00084; sushi; 1.
CC DR Pfam: PF00193; Xlink; 2.
CC DR PRINTS: PR01265; LINKMODULE.
CC DR PRINTS: PR00356; ANTIFREEZE1.
CC DR ProDom: PD000918; Link; 2.
CC DR SMART: SM00032; CCP; 1.
CC DR SMART: SM00034; CLECT; 1.
CC DR SMART: SM00179; EGF_CA; 1.
CC DR SMART: SM00409; IG; 1.
CC DR SMART: SM00445; Link; 2.
CC DR PROSITE: PS00010; ASX_HYDROXYL; 1.
CC DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
CC DR PROSITE: PS00041; C_TYPE_LECTIN_2; 1.
CC DR PROSITE: PS00022; EGF_1; 3.
CC DR PROSITE: PS01186; EGF_2; 1.
CC DR PROSITE: PS01187; EGF_CA; 1.
CC DR PROSITE: PS00835; IG_LIKE; 1.
CC DR PROSITE: PS01241; LINK; 2.
CC KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
CC EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
CC FT CHAIN 1 22
CC FT CHAIN 23 1257
CC FT CHAIN 639 1257
CC FT DOMAIN 37 157
CC FT DOMAIN 158 253
CC FT DOMAIN 259 355
CC FT DOMAIN 949 985
CC FT DOMAIN 987 1023
CC FT DOMAIN 1025 1154
CC FT DOMAIN 1155 1213
CC FT DISULFID 58 139
CC FT DISULFID 181 252
CC FT DISULFID 205 226
CC FT DISULFID 279 354
CC FT DISULFID 303 324
CC FT DISULFID 953 964

FT DISULFID 958 973 BY SIMILARITY.
FT DISULFID 975 984 BY SIMILARITY.
FT DISULFID 1029 1040 BY SIMILARITY.
FT DISULFID 1057 1149 BY SIMILARITY.
FT DISULFID 1125 1141 BY SIMILARITY.
FT DISULFID 1156 1199 BY SIMILARITY.
FT DISULFID 1185 1212 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 944 944 O-LINKED (XYL. . .) (CHONDROITIN SULFATE).
FT CARBOHYD 967 967 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1164 1164 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1257 AA; 135544 MW; 992B33DCPA19E1B CRC64;
Query Match 15.78; Score 174; DB 1; Length 1257;
Best Local Similarity 30.8%; Pred. No. 1.4e-07;
Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;
QY 15 QPCYKVIYFHDTSRRLLNFEAEACRRDGGQLVSISEDEQKLIKFIEIENLLPSDGDW 74
DB 1037 QHCYR--YF---ARRAWEDAERDCRRRAGHLTSVHSPEHKFINSF-----CHENSW 1085
QY 75 IGLRRREEKQSNSTACQDLYAWTDGSIQSPRNMYVDEPS---CGSEVGVVMYHQPSPAPAG 131
DB 1086 IGLNDRTVRD-----FQWTDNTGLQYENWREKQPDNPFAGGDCVVMVAHENG--- 1134
QY 132 IGGPYMFQWDDRCNMKNFICK 154
DB 1135 -----RWNDVPCNYNLPYVCK 1150
RESULT 8
PGCV_RAT STANDARD; PRT; 2738 AA.
AC Q9ERB4; O08592; O88564; Q9R1K4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP) (Fragments).
GN CPBG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_FaxID=10116;
OX [1]
RN [1]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.
RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wight T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [2]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=98308094; PubMed=9642104;
RA Milev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,
RA Margolis R.K., Margolis R.U.;
RT "Differential regulation of expression of hyaluronan-binding
RT proteoglycans in developing brain: aggrecan, versican, neurocan, and
RT brevican.";
RL Biochem. Biophys. Res. Commun. 247:207-212(1998).
RN [3]
RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).
RC TISSUE=Kidney;
RX MEDLINE=98094159; PubMed=9434070;
RA Pyke C., Kristensen P., Oetegaard P.B., Oturai P.S., Romer J.;
RT "Proteoglycan expression in the normal rat kidney.";
RL Nephron 77:461-470(1997).

[4]
 RN SEQUENCE OF 2535-2738 FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE=Lung;
 RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;
 RT "Molecular cloning and characterization of two developmentally
 regulated genes in rat lung."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May play a role in intercellular signaling and in
 connecting cells with the extracellular matrix. May take part in
 the regulation of cell motility, growth and differentiation. Binds
 hyaluronin acid.
 CC -!- SUBUNIT: Interacts with FBLN1 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Additional isoforms seem to exist;
 CC Name=V0;
 CC IsoId=Q9ERB4-1; Sequence=Displayed;
 CC Name=V3;
 CC IsoId=Q9ERB4-2; Sequence=VSP_003091;
 CC Name=Vint;
 CC IsoId=Q9ERB4-3; Sequence=VSP_003092;
 CC -!- TISSUE SPECIFICITY: In kidney is expressed in the papillary area,
 but not in glomeruli.
 CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
 (By similarity).
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF062402; AAC40166.1; -
 DR EMBL; U75306; AAB51125.1; -
 DR EMBL; AF084544; AAD48544.1; -
 DR EMBL; AF072892; AAC26116.1; -
 DR EMBL; AY007691; AAG16631.1; -
 DR HSP; P01132; 1EPG.
 DR InterPro; IPR000152; Asx hydroxyl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR PRINTS; PR01265; LINKMODULE.
 DR PRINTS; PR00356; ANTIFREEZEII.
 DR P-odom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLCST; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
 DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS50835; IG LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;

KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2738 VERSICAN CORE PROTEIN.
 FT NON CONS 348 349
 FT DOMAIN 21 146 IG-LIKE V-TYPE.
 FT DOMAIN 167 244 LINK 1.
 FT DOMAIN 265 346 LINK 2.
 FT DOMAIN <349 695 GAG-ALPHA
 (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
 FT DOMAIN 596 2431 GAG-BETA.
 FT DOMAIN 2431 EGF-LIKE 1.
 FT DOMAIN 2469 2505 EGF-LIKE 2, CALCIUM-BINDING.
 FT DOMAIN 2518 2632 C-TYPE LECTIN.
 FT DOMAIN 2637 2695 SUSHI.
 FT DISULFID 44 130 BY SIMILARITY.
 FT DISULFID 172 243 BY SIMILARITY.
 FT DISULFID 196 217 BY SIMILARITY.
 FT DISULFID 270 345 BY SIMILARITY.
 FT DISULFID 294 315 BY SIMILARITY.
 FT DISULFID 2435 2446 BY SIMILARITY.
 FT DISULFID 2440 2455 BY SIMILARITY.
 FT DISULFID 2457 2466 BY SIMILARITY.
 FT DISULFID 2473 2484 BY SIMILARITY.
 FT DISULFID 2478 2493 BY SIMILARITY.
 FT DISULFID 2495 2504 BY SIMILARITY.
 FT DISULFID 2511 2522 BY SIMILARITY.
 FT DISULFID 2539 2631 BY SIMILARITY.
 FT DISULFID 2607 2623 BY SIMILARITY.
 FT DISULFID 2638 2681 BY SIMILARITY.
 FT DISULFID 2667 2694 BY SIMILARITY.
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 692 692 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 758 758 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 805 805 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1257 1257 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1435 1435 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1633 1633 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1660 1660 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1684 1684 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1738 1738 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1848 1848 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2004 2004 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2409 2409 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2711 2711 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2721 2721 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 349 2431 Missing (in isoform V3).
 FT VARSPLIC 2697 2738 /FTid=VSP_003091
 FT VARSPLIC 2697 2738 PSAYQRYSKYLKNSVSKNSINTSKHEHRSRRWOETR
 R -> RKWSFRKNGQPCFNKY (in isoform Vint).
 FT CONFLICT 2535 2539 /FTid=VSP_003092.
 FT SEQUENCE 2738 AA; 300004 MW; 12CA626DS8B8DC6A CRC64;
 AEREC -> NSARG (IN REF. 4).
 Query Match 15.7%; Score 174; DB 1; Length 2738;
 Best Local Similarity 28.5%; Pred. No. 3.8e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
 Qy 15 QPCYKVIYFHDTSRRLNFEAKERDRDGGQLVSIIEDEOKLIEPIENLLPSDQDF- 73
 Db 2519 QGQCYK--YF---AHRRTWDAARECKLOQAHLTSLSHSEQMFPNRY-----GHDYQ 2566
 Qy 74 WGLRRREKQSNSTACQDLYAWTDGSGISQFRNMYVDEP-----SCGSEVCVVMYHQPSAP 129
 Db 2567 WIGL-----NDKMFEDHFRWTDGSLQYENWRPNQDPSFPFSGEDCVIWHENG-- 2616
 Qy 130 AGTGGPYMFQWDDRCMKGNFKICKYS-----DEKPAVPSREAGE 170
 Db 2617 -----QWNDVPCNYHLTYTCKKGTACVAGQPPVVENAKTFGK 2652

RESULT 9
PGCV MOUSE
ID PGCV MOUSE STANDARD; PRT; 3358 AA.
AC Q62059; Q62058; Q9CUU0;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Chondroitin core protein precursor (large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
GN CSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS V0; V1 AND V2).
RC STRAIN=C57BL/6; and Swiss Webster; TISSUE=Brain;
RX MEDLINE=95122551; PubMed=7822336;
RA Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;
RT "Multiple forms of mouse PG-M, a large chondroitin sulfate
RT proteoglycan generated by alternative splicing.";
RL J. Biol. Chem. 270:958-965(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM V3).
RC STRAIN=C57BL/6;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
RL J. Biol. Chem. 270:3914-3918(1995).
RN [3]
RP SEQUENCE OF 1-1692 FROM N.A. (ISOFORM V1).
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinesawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Ando S., Fukuda S.,
RA Azawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleichmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature, 409:685-690(2001).
RN [4]
RP INTERACTION WITH FBLN1.
RX PubMed=10400671;
RA Asperberg A., Adam S., Kostka G., Timpl R., Heinegaard D.;
RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and
RT versican.";
RL J. Biol. Chem. 274:20444-20449(1999).
CC -1- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -1- SUBUNIT: Interacts with FBLN1.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=Q62059-1; Sequence=Displayed;
CC

Name=V1;
IsoId=Q62059-2; Sequence=VSP_003087, VSP_003088;
Name=V2;
IsoId=Q62059-3; Sequence=VSP_003089;
Name=V3;
IsoId=Q62059-4; Sequence=VSP_003087, VSP_003090;
TISSUE SPECIFICITY: V2 is found only in brain.
-1- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-1- SIMILARITY: Contains 2 link domains.
-1- SIMILARITY: Contains 2 EGF-like domains.
-1- SIMILARITY: Contains 1 C-type lectin family domain.
-1- SIMILARITY: Contains 1 Sushi (SCR) domain.
-1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL; D16263; BAA03796.1; -;
DR EMBL; D28599; -; NOT ANNOTATED_CDS.
DR EMBL; D32040; BAA06802.1; -;
DR EMBL; AK014525; BAB29411.1; -;
DR HSSP; P01132; IEPG.
DR MGD; MGI:102889; Cepg2.
DR InterPro; IPR000152; Aex_hydroxyl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR005538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_Ca; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE-LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_Ca; 1.
DR PROSITE; PS08335; IG_LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; EGF-like domain; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3358 VERSICAN CORE PROTEIN.
FT DOMAIN 21 146 IG-LIKE V-TYPE.
FT DOMAIN 167 244 LINK 1.
FT DOMAIN 265 346 LINK 2.
FT DOMAIN 348 1308 GAG-ALPHA
FT (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
FT DOMAIN 1309 3052 GAG-BETA.
FT DOMAIN 3052 3088 EGF-LIKE 1.
FT DOMAIN 3090 3126 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3139 3253 C-TYPE LECTIN.

FT DOMAIN 3258 3316 SUSHI.
 FT DISULFID 44 130 BY SIMILARITY.
 FT DISULFID 172 243 BY SIMILARITY.
 FT DISULFID 196 217 BY SIMILARITY.
 FT DISULFID 270 333 BY SIMILARITY.
 FT DISULFID 294 315 BY SIMILARITY.
 FT DISULFID 3056 3067 BY SIMILARITY.
 FT DISULFID 3061 3076 BY SIMILARITY.
 FT DISULFID 3078 3087 BY SIMILARITY.
 FT DISULFID 3094 3105 BY SIMILARITY.
 FT DISULFID 3099 3114 BY SIMILARITY.
 FT DISULFID 3116 3125 BY SIMILARITY.
 FT DISULFID 3132 3143 BY SIMILARITY.
 FT DISULFID 3160 3252 BY SIMILARITY.
 FT DISULFID 3228 3244 BY SIMILARITY.
 FT DISULFID 3259 3302 BY SIMILARITY.
 FT DISULFID 3288 3315 BY SIMILARITY.
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 951 951 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1305 1305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1372 1372 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1679 1679 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2054 2054 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2244 2244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2362 2362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2827 2827 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3030 3030 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3332 3332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3342 3342 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 348 348 P -> R (in isoform V1 and isoform V3).
 FT VARSPLIC 349 1308 /FTid=VSP 003087.
 FT VARSPLIC 1309 3052 Missing (in isoform V1).
 FT VARSPLIC 349 3052 Missing (in isoform V2).
 FT VARSPLIC 349 3052 Missing (in isoform V3).
 FT CONFLICT 126 126 A -> G (IN REF. 3).
 FT CONFLICT 348 348 MISSING (IN REF. 3).
 FT CONFLICT 1658 1658 I -> T (IN REF. 3).
 FT CONFLICT 1674 1680 TVMNSNS -> QFGIQTA (IN REF. 3).
 SQ SEQUENCE 3358 AA; 366938 MW; 071B80026BC0762D CRC64;
 Query Match 15.7%; Score 174; DB 1; Length 3358;
 Best Local Similarity 28.5%; Pred. No. 4.8e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
 QY 15 QPCYKVIYCHDTSRLNFEAKACRRDGGQLVSESEDEOKLIEFNLPSDGF- 73
 Db 3140 QGQCYK--YF---AHRITWDAARECLQAGHLTSLSHSEQMFVNRV-----GHYDQ 3187
 QY 74 WIGLRREKOSNSTACQDIYAMTDSISQFRNMYVDEP-----SCGSEVCVMYHOPSAP 129
 Db 3188 WIGL-----NDKMPEDHFRWTDGSLQVYENRPNQPSFFSAGDCVVIWHENG-- 3237
 QY 130 AGIGGPMYFOWNDRCNKNFNCKYS-----DEKPAVPSREAGE 170
 Db 3238 -----QWNVDPVCNHYLTCTCKGTACGQPPVVENAKTRGK 3273
 RESULT 10
 PGCV HUMAN
 ID PGCV_HUMAN STANDARD; PRT; 3396 AA.
 AC F13611; P20754; Q13010; Q13189; Q15123; Q9UNWS;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Versican core protein precursor (Large fibroblast proteoglycan)

DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
 DE hyaluronate-binding protein) (GHAP).
 GN CSPG2
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM V0).
 RX MEDLINE=95105188; PubMed=7528742;
 RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
 RT "Characterization of the complete genomic structure of the human
 RT versican gene and functional analysis of its promoter.";
 RL J. Biol. Chem. 269:32999-33008(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM V1).
 RX TISSUE=Placenta;
 RA Zimmermann D.R., Ruoslahti E.;
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";
 RL EMBO J. 8:2975-2981(1989).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM V2).
 RX TISSUE=Glial tumor;
 RA Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "A novel glycosaminoglycan attachment domain identified in two
 RT alternative splice variants of human versican.";
 RL J. Biol. Chem. 269:32992-32998(1994).
 RN [4]
 RP SEQUENCE OF 2711-3396 FROM N.A.
 RX TISSUE=Lung fibroblast;
 RA MEDLINE=88007514; PubMed=2820964;
 RT Krusius T., Gehlsen K.R., Ruoslahti E.;
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
 RT lectin-like and growth factor-like sequences.";
 RL J. Biol. Chem. 262:13120-13125(1987).
 RN [5]
 RP SEQUENCE OF 251-347 FROM N.A.
 RX MEDLINE=93122792; PubMed=1478664;
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
 RA McPherson J.D.;
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
 RT human chromosome 5 (5q12-5q14).";
 RL Genomics 14:845-851(1992).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM V3).
 RX TISSUE=Brain;
 RA MEDLINE=95181355; PubMed=7876137;
 RT Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
 RT without a chondroitin sulfate attachment in region in mouse and human
 RT tissues.";
 RL J. Biol. Chem. 270:3914-3918(1995).
 RN [7]
 RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
 RX TISSUE=Aortic smooth muscle;
 RA MEDLINE=93327053; PubMed=10397680;
 RT Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
 RT Wright T.N.;
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
 RN [8]
 RP PARTIAL SEQUENCE.
 RX TISSUE=Brain;
 RA MEDLINE=89174663; PubMed=2466833;
 RT Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
 RT "Isolation and partial characterization of a glial
 RT hyaluronate-binding protein.";
 RL J. Biol. Chem. 264:5981-5987(1989).
 RN [9]
 RP TISSUE SPECIFICITY OF ISOFORMS.
 RX MEDLINE=96213482; PubMed=8627343;

RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "Differential expression of versican isoforms in brain tumors.";
 RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
 CC -I- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC
 CC -I- SUBUNIT: Interacts with FBLN1 (By similarity).
 CC
 CC -I- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Comment=Additional isoforms seem to exist;
 CC Name=V0;
 CC IsoId=P13611-1; Sequence=Displayed;
 CC
 CC Name=V1;
 CC IsoId=P13611-2; Sequence=VSP_003082, VSP_003083;
 CC
 CC Name=V2;
 CC IsoId=P13611-3; Sequence=VSP_003084;
 CC
 CC Name=V3;
 CC IsoId=P13611-4; Sequence=VSP_003082, VSP_003085;
 CC
 CC Name=Vint;
 CC IsoId=P13611-5; Sequence=VSP_003086;
 CC
 CC -I- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
 CC in normal brain, gliomas, medulloblastomas, schwannomas,
 CC neurofibromas, and meningiomas; v2 is restricted to normal brain
 CC and gliomas; v3 is found in all these tissues except
 CC medulloblastomas.
 CC
 CC -I- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
 CC
 CC -I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC
 CC -I- SIMILARITY: Contains 2 link domains.
 CC
 CC -I- SIMILARITY: Contains 2 EGF-like domains.
 CC
 CC -I- SIMILARITY: Contains 1 C-type lectin family domain.
 CC
 CC -I- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC
 CC -I- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; U16306; AAA65018.1; -;
 CC EMBL; X15998; CAA34128.1; -;
 CC EMBL; S52488; AAB24878.1; -;
 CC EMBL; U25555; AAA67565.1; -;
 CC EMBL; D32039; BAA06801.1; -;
 CC EMBL; J02814; AAA36437.1; -;
 CC EMBL; AF084545; AAD48545.1; -;
 CC PIR; S06014; A60979.
 CC HSP; P01132; IEGF.
 CC Genew; HGNC:2464; CSPG2.
 CC
 CC MTM; I18661; -;
 CC GO; GO:0005578; C:extracellular matrix; TAS.
 CC DR GO; GO:0005204; F:chondroitin sulfate proteoglycan; TAS.
 CC DR GO; GO:0005540; F:hyaluronic acid binding activity; TAS.
 CC DR GO; GO:0008037; P:cell recognition; TAS.
 CC DR GO; GO:0007275; P:development; TAS.
 CC DR InterPro; IPR000152; Asx_hydroxyl.
 CC DR InterPro; IPR000742; EGF_2.
 CC DR InterPro; IPR001881; EGF_Ca.
 CC DR InterPro; IPR006209; EGF like.
 CC DR InterPro; IPR007110; Ig-like.
 CC DR InterPro; IPR003599; Ig.
 CC DR InterPro; IPR003006; Ig_MHC.
 CC DR InterPro; IPR001304; Lectin_C.
 CC DR InterPro; IPR000538; Link.
 CC DR InterPro; IPR000436; Sushi_SCR_CCP.
 CC DR Pfam; PF00008; EGF; 2.
 CC DR Pfam; PF00059; lectin_C; 1.
 CC DR Pfam; PF00084; sushi; 1.
 CC DR Pfam; PF00193; Xlink; 2.

DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
 DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS08035; IG LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT CHAIN 1 20
 FT SIGNAL 1 20
 FT DOMAIN 21 3396
 FT DOMAIN 21 146
 FT DOMAIN 167 244
 FT DOMAIN 265 346
 FT DOMAIN 348 1335
 FT DOMAIN 1336 3089
 FT DOMAIN 3089 3125
 FT DOMAIN 3127 3163
 FT DOMAIN 3176 3290
 FT DOMAIN 3295 3353
 FT DISULFID 44 130
 FT DISULFID 172 243
 FT DISULFID 196 217
 FT DISULFID 270 345
 FT DISULFID 294 315
 FT DISULFID 3093 3104
 FT DISULFID 3098 3113
 FT DISULFID 3115 3124
 FT DISULFID 3131 3142
 FT DISULFID 3136 3151
 FT DISULFID 3153 3162
 FT DISULFID 3169 3180
 FT DISULFID 3197 3289
 FT DISULFID 3265 3281
 FT DISULFID 3296 3339
 FT DISULFID 3325 3352
 FT CARBOHYD 57 57
 FT CARBOHYD 330 330
 FT CARBOHYD 615 615
 Query Match 15.7%; Score 174; DB 1; Length 3396;
 Best Local Similarity 28.5%; Pred. No. 4.9e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
 QY 15 QRCYKVIYFHTSRLNFEAEKACRRDGGQVLVSIESEDEQKLEKFIENLLPSGDGF- 73
 Db 3177 QGQCYK--YF---AHRRTWDAERCKLQAGHLTSLSHBEQMFVNRV-----GHDYQ 3224
 QY 74 WGLRRREKQSNSTACQDLYAWTDCGISQFRWYVDEP-----SCGSEVCVMYHQPSAP 129
 Db 3225 WIGL-----NDKMFHEHFRWTDGSLTQYENRPNQDPSFSGEDCVIWHENG-- 3274
 QY 130 AGIGGPFMFQWDDRCNMKNFKICKYS----DEKPAVPSPREARGE 170
 Db 3275 -----QWDDVPCNYHLTYTCKKGTVACGGPPVVENAKTGGK 3310
 RESULT 11
 PGCV CHICK
 ID PGCV CHICK STANDARD; PRT; 3562 AA.
 AC Q90953; Q90945;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)

DR Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (pg-M).
OS CPSC2.
GN Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS VO AND V1).
RC STRAIN=White leghorn; TISSUE=limb bud;
RX MEDLINE=93300846; PubMed=8314802;
RA Shinomura T., Nishida Y., Ito K., Kimata K.;
RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
RT expressed during chondrogenesis in chick limb buds. Alternative
RT spliced multiforms of pg-M and their relationships to versican.";
RL J. Biol. Chem. 268:14461-14469(1993).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing. Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=Q90953-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=Q90953-2; Sequence=VSP_003093;
CC -!- TISSUE SPECIFICITY: Prechondrogenic condensation area of
CC developing limb buds.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X60226; CAA42787.1; -;
DR EMBL; D13542; BAA02742.1; -;
DR PIR; A47171; A47171.
DR HSP; P00740; 1EDM.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_C; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG_1.
DR SMART; SM00445; LINK; 2.

DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 27 3562
FT DOMAIN 27 143
FT DOMAIN 166 243
FT DOMAIN 264 345
FT DOMAIN 3254 3290
FT DOMAIN 3292 3328
FT DOMAIN 3341 3455
FT DOMAIN 3460 3518
FT DISULFID 44 129
FT DISULFID 171 242
FT DISULFID 195 216
FT DISULFID 269 344
FT DISULFID 293 314
FT DISULFID 3258 3269
FT DISULFID 3263 3278
FT DISULFID 3280 3289
FT DISULFID 3296 3307
FT DISULFID 3301 3316
FT DISULFID 3318 3327
FT DISULFID 3334 3345
FT DISULFID 3362 3454
FT DISULFID 3430 3446
FT DISULFID 3461 3504
FT DISULFID 3490 3517
FT CARBOHYD 163 163
FT CARBOHYD 235 235
FT CARBOHYD 329 329
FT CARBOHYD 529 529
FT CARBOHYD 709 709
FT CARBOHYD 948 948
FT CARBOHYD 1409 1409
FT CARBOHYD 1479 1479
FT CARBOHYD 1523 1523
FT CARBOHYD 1530 1530
FT CARBOHYD 1625 1625
FT CARBOHYD 1751 1751
FT CARBOHYD 1988 1988
FT CARBOHYD 2088 2088
FT CARBOHYD 2089 2089
FT CARBOHYD 2507 2507
FT CARBOHYD 2642 2642
FT CARBOHYD 2679 2679
FT CARBOHYD 2748 2748
FT CARBOHYD 2762 2762
FT CARBOHYD 3069 3069
FT CARBOHYD 3194 3194
FT CARBOHYD 3232 3232
FT CARBOHYD 3545 3545
FT VARSPLIC 485 1411
FT FTid=VSP_003093
SQ SEQUENCE 3562 AA; 388078 MW; 9BC56588C1602D2 CRC64;

Query Match 15.5%; Score 171; DB 1; Length 3562;
Best Local Similarity 28.5%; Pred. No. 9.7e-07;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

Qy 15 QRCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSEIEBOKLIEFIENLLSDGDF- 73
Db 3342 QGQCYK--YF---AHRRTWDTARECRLOGAHLTSLSHSEEQVFNRI-----GHDYQ 3389
Qy 74 WIGLRREEKQNSNACQDLVATDGTSGISQFRNYYVDEP-----SCGSEVCVVMYHQSAP 129

```

Db 3390 WIGL-----NDKMFEDFRWTDGSLQYENRPNQDSFFSAGEDCVVLIWHENG-- 3439
Qy 130 AGIGGPMFQWDDRCNKNKFNICKYS-----DEKPAVPSREAGE 170
Db 3440 -----QWNVPCNYHLTYTCKGTGACGPPVVENAKTFGK 3475

RESULT 12
CD93_RAT
ID CD93_RAT STANDARD; PRT; 643 AA.
AC Q9ET61; Q9J126;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Complement component C1q receptor precursor (Complement component 1, q
subcomponent, receptor 1) (C1qR) (C1qR(p)) (C1q/MBL/SPA receptor)
DE (CD93 antigen) (Cell surface antigen AA4).
GN C1QR1 OR CD93 OR C1QRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=PVG; TISSUE=Natural killer cells;
RX MEDLINE=20545218; PubMed=11093152;
RA Lovik G., Vaage J.T., Dissen E., Szpirer C., Ryan J.C., Rolstad B.;
RT "Characterization and molecular cloning of rat C1qR, a receptor on NK
cells.";
RL Eur. J. Immunol. 30:3355-3362(2000).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=Mistar; TISSUE=Lung;
RX MEDLINE=20507883; PubMed=10934210;
RA Dean Y.D., McGreal E.P., Akatsu H., Gaeque P.;
RT "Molecular and cellular properties of the rat AA4 antigen, a C-type
lectin-like receptor with structural homology to thrombomodulin.";
RL J. Biol. Chem. 275:34382-34392(2000).
CC -1- FUNCTION: Receptor (or element of a larger receptor complex) for
CC C1q, mannose-binding lectin (MBL2) and pulmonary surfactant
CC protein A (SPA). May mediate the enhancement of phagocytosis in
CC monocytes and macrophages upon interaction with soluble defense
CC collagens. May play a role in intercellular adhesion.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Widely expressed. Highly expressed in lung and
CC heart. Expressed at lower level in brain, thymus, liver, spleen,
CC intestine, kidney, adrenal gland, muscle and testis. Expressed on
CC circulating natural killer cells.
CC -1- PTM: N- and O-glycosylated (by similarity).
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 5 EGF-like domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF136537; AAG01572.1; -.
CC EMBL; AF160978; AAF80402.1; -.
CC HSSP; P35555; 1EMN.
CC GO; GO:0016021; C:integral to membrane; ISS.
CC GO; GO:0004872; F:receptor activity; ISS.
CC GO; GO:0016337; P:cell-cell adhesion; ISS.
CC GO; GO:0042116; P:macrophage activation; ISS.
CC GO; GO:0006909; P:phagocytosis; ISS.
CC InterPro; IPR000152; Rex hydroxyl.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF_like.

```

```

DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00008; EGF_5.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX HYDROXYL; 3.
DR PROSITE; PS00615; C-TYPE LECTIN 1; FALSE_NEG.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 3.
KW Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
KW EGF-like domain; Lectin; Glycoprotein.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 643 COMPLEMENT COMPONENT C1Q RECEPTOR.
FT DOMAIN 24 571 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 572 592 POTENTIAL.
FT DOMAIN 593 643 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 31 173 C-TYPE LECTIN.
FT DOMAIN 257 298 EGF-LIKE 1.
FT DOMAIN 299 341 EGF-LIKE 2.
FT DOMAIN 342 381 EGF-LIKE 3.
FT DOMAIN 382 423 EGF-LIKE 4.
FT DOMAIN 424 462 EGF-LIKE 5.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 268 282 BY SIMILARITY.
FT DISULFID 284 297 BY SIMILARITY.
FT DISULFID 303 314 BY SIMILARITY.
FT DISULFID 308 325 BY SIMILARITY.
FT DISULFID 327 340 BY SIMILARITY.
FT DISULFID 346 355 BY SIMILARITY.
FT DISULFID 351 364 BY SIMILARITY.
FT DISULFID 366 380 BY SIMILARITY.
FT DISULFID 386 397 BY SIMILARITY.
FT DISULFID 393 406 BY SIMILARITY.
FT DISULFID 408 422 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 433 446 BY SIMILARITY.
FT DISULFID 448 461 BY SIMILARITY.
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 417 417 E -> K (IN REF. 2).
SQ SEQUENCE 643 AA; 68781 MW; 9AEAC933AD943DB6 CRC64;

Query Match 14.9%; Score 165; DB 1; Length 643;
Best Local Similarity 25.2%; Pred. No. 4e-07;
Matches 52; Conservative 35; Mismatches 73; Indels 46; Gaps 10;

Qy 1 GRLLSGOP-----VCRGGTQRCYKVIYFHDTSRLNFEAEKACRRDGGQLVSIIESD 54
Db 15 GQLWAGAAADSEAVVCEG---TACTAHM-----GKLSAAEAQHRCNENGGLATVKSEE 66
Qy 55 EQKLEIKFTIENLL-----PSD---GDFWIGLRREKESNSTACQDLYAWT-DGSIQSPRN 106
Db 67 EARHVOEALAQLLKTKAPSETKIGKFWIGLQREKCKTYHDLPMKGFVWGGGDDTYSN 126
Qy 107 WY-VDEPSCGSEVCVVMY-----HQPSAPAGIGGPFYMFQWDDRC-----NMKNPF 151
Db 127 WYKASKSSCLSKKCVSLILDLKLRPHSLP-----KWHESPCTPDAPGNSIEGP 177
Qy 152 ICKYSDEKPAVPREAGEETELTP 177
Db 178 LCKNFKMGCSPLALGGPGQLTVTP 203

RESULT 13
PGCA_BOVIN
ID PGCA_BOVIN STANDARD; PRT; 2364 AA.
AC P13608; P79117; Q28159;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggreacan core protein precursor (Cartilage-specific proteoglycan core
protein) (CSPCP).

```


FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 509 580 BY SIMILARITY.
FT DISULFID 533 554 BY SIMILARITY.
FT DISULFID 607 682 BY SIMILARITY.
FT DISULFID 631 652 BY SIMILARITY.
FT DISULFID 2117 2128 BY SIMILARITY.
FT DISULFID 2182 2274 BY SIMILARITY.
FT DISULFID 2250 2266 BY SIMILARITY.
FT DISULFID 2281 2324 BY SIMILARITY.
FT DISULFID 2310 2337 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 2114 2150 Missing (in isoform 2).
FT SQ SEQUENCE 2364 AA; 246359 MW; 6FF83763420C3D4C CRC64;
/FTID=VSP_003072.
Query Match 14.3%; Score 158.5; DB 1; Length 2364;
Best Local Similarity 26.8%; Pred. No. 7.6e-06;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;
Qy 7 QPVCGRG---TQPCYKVIYFHDTSRRLNFEAKEACRRGGQGLVSESEDEQKLIKFI 63
Db 2151 QKLCEGWTKFGHCYR--HPFD---RATWDAESQCRKQKQSHLSIVTPEQ----EFV 2201
Qy 64 ENLLPSDQGF-WIGLRREEKQSNSTACDLYAYTDGSIQFRWYVDEP-----SCGSEV 118
Db 2202 NN---NAQDYQWIGL-----NDKITEGDFRWSGDHSLQFNWRPNQPNFFATGBCD 2250
Qy 119 CVMVYHOPSAFAGIGPWFOMNDRCNMKNFKICKYS----DKPAVPSREAGEETE 173
Db 2251 VMIWHEKG-----BWNVPNCYQLPFTCKGTACGPPVVEHARIFGQKDD 2298
RESULT 14
PGCA HUMAN STANDARD; PRT: 2415 AA.
AC P16112; Q13650;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP) (Chondroitin sulfate proteoglycan core protein 1).
GN AGC1 OR CSPG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Chondrocytes;
RX MEDLINE=91093289; PubMed=1985970;
RA Doeg K.J., Sasaki M., Kimura T., Yamada Y.;
RT "Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms.";
RL J. Biol. Chem. 266:1894-1902(1991).
RN [2]
RP SEQUENCE OF 1778-2415 FROM N.A. (ISOFORM 2).
RC TISSUE=Chondrocytes;
RA Dudhia J., Hardingham T.E.;
RN Submitted (JAN-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1936-2415 FROM N.A. (ISOFORM 1).
RX MEDLINE=89380154; PubMed=2789216;
RA Baldwin C.T., Reginato A.M., Prockop D.J.;
RT "A new epidermal growth factor-like domain in the human core protein for the large cartilage-specific proteoglycan. Evidence for alternative splicing of the domain.";
RL J. Biol. Chem. 264:15747-15750(1989).

RN [4]
RP SEQUENCE OF 764-864 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=95128522; PubMed=7827755;
RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
RT "Length variation in the keratan sulfate domain of mammalian aggrecan.";
RL Matrix Biol. 14:323-328(1994).
CC -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION.
CC -!- SUBUNIT: Interacts with FBLN1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P16112-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P16112-2; Sequence=VSP_003074;
CC Name=3;
CC IsoId=P16112-3; Sequence=VSP_003074, VSP_003075;
CC -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 4 link domains.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M55172; AAA62824.1; -;
CC EMBL; J05062; AAA35726.1; -;
CC EMBL; X17406; CAA35463.1; -;
CC EMBL; S74659; AAC60643.2; -;
CC DR PIR; A39086; A39086.
CC DR HSP; P98066; ITSG.
CC DR Genew; HGNC:319; AGC1.
CC MW: 155760; -;
CC GO: 0005204; P:chondroitin sulfate proteoglycan; TAS.
CC InterPro; IPR002353; Antifreeze1.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR003324; SGXXSG.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00047; Ig; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF02339; SGXXSG; 71.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 4.

DR PRINTS; PRO1265; LINKMODULE.
 DR PRINTS; PRO0356; ANTIFREEZEII.
 DR ProDom; PRO00318; Link; 4.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 4.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
 DR PROSITE; PS00441; C-TYPE LECTIN 2; 1.
 DR PROSITE; PS00022; EGF 1; 1.
 DR PROSITE; PS01186; EGF 2; 1.
 DR PROSITE; PS00835; IG LIKE; 1.
 DR PROSITE; PS00290; IG MHC; 1.
 DR PROSITE; PS01241; LINK; 3.
 DR Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW EGF-like domain; Alternating Repeat; Immunoglobulin domain.
 FT SIGNAL 1 19
 FT CHAIN 20 2415
 FT DOMAIN 34 147
 FT DOMAIN 170 247
 FT DOMAIN 268 349
 FT DOMAIN 495 572
 FT DOMAIN 593 673
 FT DOMAIN 2164 2199
 FT DOMAIN 2201 2327
 FT DOMAIN 2331 2389
 FT DOMAIN 48 141
 FT DOMAIN 152 247
 FT DOMAIN 253 349
 FT DOMAIN 477 571
 FT DOMAIN 578 672
 FT DOMAIN 676 848
 FT DOMAIN 772 844
 FT DOMAIN 851 1497
 FT DOMAIN 941 1497
 FT DOMAIN 1498 2162
 FT DOMAIN 2163 2415
 FT DISULFID 51 133
 FT DISULFID 175 246
 FT DISULFID 199 220
 FT DISULFID 273 348
 FT DISULFID 297 318
 FT DISULFID 500 571
 FT DISULFID 524 545
 FT DISULFID 598 672
 FT DISULFID 621 642
 FT DISULFID 2168 2178
 FT DISULFID 2173 2187
 FT DISULFID 2189 2198
 FT DISULFID 2205 2216
 FT DISULFID 2303 2325
 FT DISULFID 2301 2317
 FT DISULFID 2332 2375
 FT DISULFID 2361 2388
 FT CARBOHYD 126 126
 FT CARBOHYD 239 239
 FT CARBOHYD 333 333
 FT CARBOHYD 387 387
 FT CARBOHYD 434 434
 FT CARBOHYD 602 602
 FT CARBOHYD 657 657
 FT CARBOHYD 737 737
 FT CARBOHYD 1898 1898
 FT VARSPLIC 2163 2200
 FT VARSPLIC 2330 2390
 FT CONFLICT 766 766
 FT CONFLICT 847 847
 FT CONFLICT 1928 1928
 FT CONFLICT 1964 1964
 FT CONFLICT 2070 2070

FT CONFLICT 2391 2391 A -> P (IN REF. 2 AND 3).
 SQ SEQUENCE 2415 AA; 250191 MW; 1288937ELB98C6B6 CRC64;
 Query Match 14.3%; Score 158.5; DB 1; Length 2415;
 Best Local Similarity 29.1%; Pred. No. 7.8e-06;
 Matches 52; Conservative 22; Mismatches 62; Indels 43; Gaps 11;
 QY 7 QVCRGG--TORPCYKVIYFHDTSRLNFEAKACRRDGGQLVISEDQKLIETFI 63
 DB 2202 QEVCEGNKYQHGYR--HFPD---RETWDAERRCREQQHLSIVTPSEQ----EFV 2252
 QY 64 ENLLPSDGDGF-WIGLRRREERQSNSTACQDLYAMTDGSIQFRNMYVDPS---CGSEVC 119
 DB 2253 NN---NAQDYQWIGL-----NDRTIEGFRMSDGHMQRPNQDNFFRAGEDC 2301
 QY 120 VVM--YHQPSAPAGIGGYPWFQNDRCNMKNFICKYS----DEKPAVPSRAEGETE 173
 DB 2302 VVMWHEKG-----EWNDVPCNVHLPFTCKGTACGEPVVEHARTFGOKKD 2349
 RESULT 15
 PGCA CANFA STANDARD; PRT; 2333 AA.
 ID PGCA CANFA Q28343; Q28310;
 AC Q28343; Q28310;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
 DE protein) (CSPCP).
 GN AGC1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Glant T.T., Adams M.E., Kwok S.X.F., Huang D., Fuelleop C.;
 RT "Complete coding sequence and deduced amino acid sequence of aggrecan
 RT of canine cartilage."
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 774-833 FROM N.A.
 RC TISSUE=Cartilage;
 EX MEDLINE=95128522; PubMed=7827755;
 RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
 RT "Length variation in the keratan sulfate domain of mammalian
 RT aggrecan."
 RL Matrix Biol. 14:323-328(1994).
 RN [3]
 RP SEQUENCE OF 1830-2333 FROM N.A.
 RA Adams M.E., Kwok S.X.F., Huang D., Glant T.T., Fullop C.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 2082-2118 FROM N.A.
 RC TISSUE=Cartilage;
 EX MEDLINE=93352525; PubMed=8349621;
 RA Fuelleop C., Walcz E., Vallyon M., Glant T.T.;
 RT "Expression of alternatively spliced epidermal growth factor-like
 RT domains in aggrecans of different species. Evidence for a novel
 RT module."
 RL J. Biol. Chem. 268:17377-17383(1993).
 CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
 CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
 CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
 CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
 CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
 CC similarity).
 CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
 CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
 CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
 CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
 CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:06:34 ; Search time 9.91785 Seconds
(without alignments)
1978.090 Million cell updates/sec

Title: US-09-887-855-2_COPY_24_227

Perfect score: 1106
Sequence: 1 GRLLSGQPVCRGGTQPCYK.....EEDAKTTFKESREAAALNLAY 204

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76.*
1: Pirl.*
2: Pirl2.*
3: Pirl3.*
4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 182 | 16.5 | 1456 | 1 A36563 | mannose receptor p |
| 2 | 178.5 | 16.1 | 1455 | 1 A48925 | mannose receptor p |
| 3 | 177 | 16.0 | 1268 | 2 S52781 | neurocan - mouse |
| 4 | 174.5 | 15.8 | 1643 | 2 T14274 | versican precursor |
| 5 | 174.5 | 15.8 | 3381 | 2 T42389 | versican precursor |
| 6 | 174 | 15.7 | 1257 | 2 S28784 | neurocan precursor |
| 7 | 174 | 15.7 | 2397 | 1 A55535 | versican precursor |
| 8 | 174 | 15.7 | 2409 | 1 A60979 | versican precursor |
| 9 | 171 | 15.5 | 3562 | 2 A47171 | chondroitin sulfat |
| 10 | 170.5 | 15.4 | 1479 | 2 T42710 | mannose receptor, |
| 11 | 158.5 | 14.3 | 1340 | 2 A39808 | proteoglycan core |
| 12 | 158.5 | 14.3 | 2327 | 2 T42630 | aggreccan - bovine |
| 13 | 158.5 | 14.3 | 2415 | 1 A39086 | aggreccan precursor |
| 14 | 154.5 | 14.0 | 612 | 2 B42755 | E-selectin precurs |
| 15 | 153.5 | 13.9 | 2124 | 2 A28452 | proteoglycan core |
| 16 | 152 | 13.7 | 912 | 2 A54423 | brevician precursor |
| 17 | 149.5 | 13.5 | 459 | 2 T24425 | brevician precursor |
| 18 | 149 | 13.5 | 321 | 1 LNHUP6 | IgE Fc receptor II |
| 19 | 148.5 | 13.4 | 330 | 2 T46256 | hypothetical prote |
| 20 | 148 | 13.4 | 253 | 2 E89130 | brevician - human (|
| 21 | 147.5 | 13.3 | 2132 | 1 A55182 | protein F52E1.2 [i |
| 22 | 147 | 13.3 | 883 | 2 S57653 | aggreccan precursor |
| 23 | 146.5 | 13.2 | 162 | 1 LNRCS | brevician precursor |
| 24 | 146 | 13.2 | 2109 | 1 I50421 | lectin BRA3-1 prec |
| 25 | 145.5 | 13.2 | 742 | 2 JC7595 | aggreccan precursor |
| 26 | 145 | 13.1 | 883 | 2 S49126 | scavenger receptor |
| 27 | 144.5 | 13.1 | 173 | 2 S10548 | brevician precursor |
| 28 | 144.5 | 13.1 | 372 | 2 S23936 | lectin - barnacle |
| 29 | 144.5 | 13.1 | 404 | 2 A46274 | I-selectin precurs |
| | | | | | HIV gp120-binding |

| | | | | | |
|----|-------|------|------|----------|--------------------|
| 30 | 143.5 | 13.0 | 129 | 2 JC4329 | coagulation factor |
| 31 | 143.5 | 13.0 | 372 | 1 A32375 | L-selectin precurs |
| 32 | 143.5 | 13.0 | 463 | 2 T26655 | hypothetical prote |
| 33 | 142.5 | 12.9 | 131 | 2 JC5058 | bitiscetin alpha c |
| 34 | 142 | 12.8 | 1487 | 2 S48719 | phospholipase-A(2) |
| 35 | 141.5 | 12.8 | 331 | 1 LNMSE | IgE Fc receptor, 1 |
| 36 | 140.5 | 12.7 | 162 | 1 LNRCS | lectin BRA3-2 prec |
| 37 | 140 | 12.7 | 370 | 2 S22124 | L-selectin precurs |
| 38 | 139 | 12.6 | 248 | 1 LNHUP6 | pulmonary surfacta |
| 39 | 139 | 12.6 | 248 | 1 LNHUP6 | pulmonary surfacta |
| 40 | 139 | 12.6 | 248 | 1 LNHUP6 | pulmonary surfacta |
| 41 | 139 | 12.6 | 283 | 1 LNFHLS | lectin precursor - |
| 42 | 138.5 | 12.5 | 152 | 2 JC4690 | coagulation factor |
| 43 | 138.5 | 12.5 | 202 | 2 JC4031 | tetraneectin precu |
| 44 | 138 | 12.5 | 280 | 2 T29200 | hypothetical prote |
| 45 | 137.5 | 12.4 | 309 | 1 S34198 | IgE Fc receptor II |

ALIGNMENTS

RESULT 1

A36563

mannose receptor precursor - human

C/Species: Homo sapiens (man)

C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999
C/Accession: A36563; A60926; A44255; B44255; C44255; D44255; E44255; F44255; G44255; H44255; I44255; J44255; K44255; L44255; M44255; N44255; O44255; P44255; Q44255; R44255; S44255; T44255; U44255; V44255; W44255; X44255; Y44255; Z44255; AA44255; AB44255; AC44255; AD44255; AE44255; AF44255; AG44255; AH44255; AI44255; AJ44255; AK44255; AL44255; AM44255; AN44255; AO44255; AP44255; AQ44255; AR44255; AS44255; AT44255; AU44255; AV44255; AW44255; AX44255; AY44255; AZ44255; BA44255; BB44255; BC44255; BD44255; BE44255; BF44255; BG44255; BH44255; BI44255; BJ44255; BK44255; BL44255; BM44255; BN44255; BO44255; BP44255; BQ44255; BR44255; BS44255; BT44255; BU44255; BV44255; BW44255; BX44255; BY44255; BZ44255; CA44255; CB44255; CC44255; CD44255; CE44255; CF44255; CG44255; CH44255; CI44255; CJ44255; CK44255; CL44255; CM44255; CN44255; CO44255; CP44255; CQ44255; CR44255; CS44255; CT44255; CU44255; CV44255; CW44255; CX44255; CY44255; CZ44255; DA44255; DB44255; DC44255; DD44255; DE44255; DF44255; DG44255; DH44255; DI44255; DJ44255; DK44255; DL44255; DM44255; DN44255; DO44255; DP44255; DQ44255; DR44255; DS44255; DT44255; DU44255; DV44255; DW44255; DX44255; DY44255; DZ44255; EA44255; EB44255; EC44255; ED44255; EE44255; EF44255; EG44255; EH44255; EI44255; EJ44255; EK44255; EL44255; EM44255; EN44255; EO44255; EP44255; EQ44255; ER44255; ES44255; ET44255; EU44255; EV44255; EW44255; EX44255; EY44255; EZ44255; FA44255; FB44255; FC44255; FD44255; FE44255; FF44255; FG44255; FH44255; FI44255; FJ44255; FK44255; FL44255; FM44255; FN44255; FO44255; FP44255; FQ44255; FR44255; FS44255; FT44255; FU44255; FV44255; FW44255; FX44255; FY44255; FZ44255; GA44255; GB44255; GC44255; GD44255; GE44255; GF44255; GH44255; GI44255; GJ44255; GK44255; GL44255; GM44255; GN44255; GO44255; GP44255; GQ44255; GR44255; GS44255; GT44255; GU44255; GV44255; GW44255; GX44255; GY44255; GZ44255; HA44255; HB44255; HC44255; HD44255; HE44255; HF44255; HG44255; HH44255; HI44255; HJ44255; HK44255; HL44255; HM44255; HN44255; HO44255; HP44255; HQ44255; HR44255; HS44255; HT44255; HU44255; HV44255; HW44255; HX44255; HY44255; HZ44255; IA44255; IB44255; IC44255; ID44255; IE44255; IF44255; IG44255; IH44255; II44255; IJ44255; IK44255; IL44255; IM44255; IN44255; IO44255; IP44255; IQ44255; IR44255; IS44255; IT44255; IU44255; IV44255; IW44255; IX44255; IY44255; IZ44255; JA44255; JB44255; JC44255; JD44255; JE44255; JF44255; JG44255; JH44255; JI44255; JJ44255; JK44255; JL44255; JM44255; JN44255; JO44255; JP44255; JQ44255; JR44255; JS44255; JT44255; JU44255; JV44255; JW44255; JX44255; JY44255; JZ44255; KA44255; KB44255; KC44255; KD44255; KE44255; KF44255; KG44255; KH44255; KI44255; KL44255; KM44255; KN44255; KO44255; KP44255; KQ44255; KR44255; KS44255; KT44255; KU44255; KV44255; KW44255; KX44255; KY44255; KZ44255; LA44255; LB44255; LC44255; LD44255; LE44255; LF44255; LG44255; LH44255; LI44255; LJ44255; LK44255; LL44255; LM44255; LN44255; LO44255; LP44255; LQ44255; LR44255; LS44255; LT44255; LU44255; LV44255; LW44255; LX44255; LY44255; LZ44255; MA44255; MB44255; MC44255; MD44255; ME44255; MF44255; MG44255; MH44255; MI44255; MJ44255; MK44255; ML44255; MM44255; MN44255; MO44255; MP44255; MQ44255; MR44255; MS44255; MT44255; MU44255; MV44255; MW44255; MX44255; MY44255; MZ44255; NA44255; NB44255; NC44255; ND44255; NE44255; NF44255; NG44255; NH44255; NI44255; NJ44255; NK44255; NL44255; NM44255; NO44255; NP44255; NQ44255; NR44255; NS44255; NT44255; NU44255; NV44255; NW44255; NX44255; NY44255; NZ44255; OA44255; OB44255; OC44255; OD44255; OE44255; OF44255; OG44255; OH44255; OI44255; OJ44255; OK44255; OL44255; OM44255; ON44255; OO44255; OP44255; OQ44255; OR44255; OS44255; OT44255; OU44255; OV44255; OW44255; OX44255; OY44255; OZ44255; PA44255; PB44255; PC44255; PD44255; PE44255; PF44255; PG44255; PH44255; PI44255; PJ44255; PK44255; PL44255; PM44255; PN44255; PO44255; PP44255; PQ44255; PR44255; PS44255; PT44255; PU44255; PV44255; PW44255; PX44255; PY44255; PZ44255; QA44255; QB44255; QC44255; QD44255; QE44255; QF44255; QG44255; QH44255; QI44255; QJ44255; QK44255; QL44255; QM44255; QN44255; QO44255; QP44255; QQ44255; QR44255; QS44255; QT44255; QU44255; QV44255; QW44255; QX44255; QY44255; QZ44255; RA44255; RB44255; RC44255; RD44255; RE44255; RF44255; RG44255; RH44255; RI44255; RJ44255; RK44255; RL44255; RM44255; RN44255; RO44255; RP44255; RQ44255; RR44255; RS44255; RT44255; RU44255; RV44255; RW44255; RX44255; RY44255; RZ44255; SA44255; SB44255; SC44255; SD44255; SE44255; SF44255; SG44255; SH44255; SI44255; SJ44255; SK44255; SL44255; SM44255; SN44255; SO44255; SP44255; SQ44255; SR44255; SS44255; ST44255; SU44255; SV44255; SW44255; SX44255; SY44255; SZ44255; TA44255; TB44255; TC44255; TD44255; TE44255; TF44255; TG44255; TH44255; TI44255; TJ44255; TK44255; TL44255; TM44255; TN44255; TO44255; TP44255; TQ44255; TR44255; TS44255; TT44255; TU44255; TV44255; TW44255; TX44255; TY44255; TZ44255; UA44255; UB44255; UC44255; UD44255; UE44255; UF44255; UG44255; UH44255; UI44255; UJ44255; UK44255; UL44255; UM44255; UN44255; UO44255; UP44255; UQ44255; UR44255; US44255; UT44255; UU44255; UV44255; UW44255; UX44255; UY44255; UZ44255; VA44255; VB44255; VC44255; VD44255; VE44255; VF44255; VG44255; VH44255; VI44255; VJ44255; VK44255; VL44255; VM44255; VN44255; VO44255; VP44255; VQ44255; VR44255; VS44255; VT44255; VU44255; VV44255; VW44255; VX44255; VY44255; VZ44255; WA44255; WB44255; WC44255; WD44255; WE44255; WF44255; WG44255; WH44255; WI44255; WJ44255; WK44255; WL44255; WM44255; WN44255; WO44255; WP44255; WQ44255; WR44255; WS44255; WT44255; WU44255; WV44255; WW44255; WX44255; WY44255; WZ44255; XA44255; XB44255; XC44255; XD44255; XE44255; XF44255; XG44255; XH44255; XI44255; XJ44255; XK44255; XL44255; XM44255; XN44255; XO44255; XP44255; XQ44255; XR44255; XS44255; XT44255; XU44255; XV44255; XW44255; XX44255; XY44255; XZ44255; YA44255; YB44255; YC44255; YD44255; YE44255; YF44255; YG44255; YH44255; YI44255; YJ44255; YK44255; YL44255; YM44255; YN44255; YO44255; YP44255; YQ44255; YR44255; YS44255; YT44255; YU44255; YV44255; YW44255; YX44255; YY44255; YZ44255; ZA44255; ZB44255; ZC44255; ZD44255; ZE44255; ZF44255; ZG44255; ZH44255; ZI44255; ZJ44255; ZK44255; ZL44255; ZM44255; ZN44255; ZO44255; ZP44255; ZQ44255; ZR44255; ZS44255; ZT44255; ZU44255; ZV44255; ZW44255; ZX44255; ZY44255; ZZ44255;

A/Title: Primary structure of the mannose receptor contains multiple motifs resembling

A/Reference number: A36563; MUID:90324192; PMID:2373685

A/Accession: A36563

A/Molecule type: mRNA

A/Residues: 1-1456 <TAY>

A/Cross-references: GB:J05550; NID:g188675; PIDN:AAA59868.1; PID:g188676

A/Note: parts of this sequence, including the amino end of the mature protein, were con

R:Ezekowitz, R.A.B.; Sastry, K.; Bailly, P.; Warner, A.

J. Exp. Med. 172, 1785-1794, 1990

A/Title: Molecular characterization of the human macrophage mannose receptor: demonst

A/Reference number: A60926; MUID:91079783; PMID:2258707

A/Accession: A60926

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-1333, 'T', 1335-1456 <EZE>

A/Cross-references: GB:X55635

A/Note: translation of the nucleotide sequence is incomplete

R:Kim, S.J.; Ruiz, N.; Bezouska, K.; Drickamer, K.

Genomics 14, 721-727, 1992

A/Title: Organization of the gene encoding the human macrophage mannose receptor (MRC1

A/Reference number: A44255; MUID:93052405; PMID:1294118

A/Accession: A44255

A/Status: preliminary; nucleic acid sequence not shown; not compared with conceptual t

A/Molecule type: DNA

A/Residues: 155-233, 'KSAI', 238-283; 346-428; 492-569; 631-714, 716-719; 783-820, 'N', 822-865

A/Note: sequence extracted from NCBI backbone (NCBIP:118415, NCBIP:118421, NCBIP:118421

C/Genetics:

A/Gene: GDB:MRC1

A/Cross-references: GDB:133759; OMIM:153618

A/Map position: 10p13-10p13

C/Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II,

C/Keywords: duplication; lectin; tandem repeat; transmembrane protein

F:1-18/Domain: signal sequence #status predicted <SIG>

F:168-209/Domain: fibronectin type II repeat homology <2FI>

F:223-340/Domain: C-type lectin homology <LCH1>

F:362-486/Domain: C-type lectin homology <LCH2>

F:945-1079/Domain: C-type lectin homology <LCH3>

Query Match 16.5%; Score 182; DB 1; Length 1456;

Best Local Similarity 25.5%; Pred. No. 8.8e-08;

Matches 52; Conservative 37; Mismatches 61; Indels 54; Gaps 8;

QY 19 YKVIYFHTSRRLNFEAKACRRDQQLVSIIEDEQKLIENLLPSDGFWGLR 78

Db 807 YKDYQYFSEKETMDNARAFKRNFGDLVSIQSESEKFLWKYV-NRNDQAQSAFYGLL 865
QY 79 RREEKOSNSTACODLYAWTDGSIQFRNMYVDPSGCS--EVCVMYTHQPSAPAGIGPY 136
Db 866 ISLDKK-----FAMMDGSKVDYVYSWATGPNFANEDENCVTWY-----SNSGF--- 908
QY 137 MFQWNDRCNMKNFICKYDEK---PAVPSREAEGEETELTPVLPEBTOE----- 185
Db 909 ---WINDINGCPYNAFICQRHNSINATVMP-----TMPSPVSCCKEGWNFYSN 954
QY 186 -----EDAKTKFKESREAAAL 200
Db 955 KCFKIFGFMEERKNQWEARKACI 978

RESULT 2
A48925
mannose receptor precursor, macrophage - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A48925; S21320; PC2245
R;Harris, N.; Super, M.; Rits, M.; Chang, G.; Ezekowitz, R.A.
Blood 80, 2363-2373, 1992
A;Title: Characterization of the murine macrophage mannose receptor: demonstration that
A;Reference number: A48925; MUID:93043353; PMID:1421407
A;Accession: A48925
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1455 <HA>
A;Experimental source: peritoneal macrophage
A;Note: sequence extracted from NCBI backbone (NCBIPI:118733)
R;Harris, N.; Super, M.; Rits, M.; Chang, G.; Ezekowitz, R.B.
submitted to the EMBL Data Library, April 1992
A;Description: Characterization of the murine macrophage mannose receptor: Demonstration
on.
A;Reference number: S21320
A;Accession: S21320
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-302, 'W', 303-1117, 'E', 1119-1455 <HA2>
R;Harris, N.; Peters, L.L.; Eicher, E.M.; Rits, M.; Rasperry, D.; Eichbaum, Q.G.; Super
Biochem. Biophys. Res. Commun. 198, 682-692, 1994
A;Title: The exon-intron structure and chromosomal localization of the mouse macrophage
A;Reference number: PC2245; MUID:94128116; PMID:8297379
A;Accession: PC2245
A;Molecule type: mRNA
A;Residues: 35-105 <HA3>
C;Genetics:
A;Gene: Mrcl
A;Map position: 2
C;Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II re
C;Keywords: membrane protein; receptor
F;168-209/Domain: fibronectin type II repeat homology <2F9>
F;361-485/Domain: C-type lectin homology <LCH1>
F;943-1077/Domain: C-type lectin homology <LCH2>

Query Match 16.1%; Score 178.5; DB 1; Length 1455;
Best Local Similarity 25.3%; Pred. No. 1.8e-07;
Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;
QY 19 YKVIYFHTSRRLNFEAEKACRRDGGQVLVSIESEDEQKLIKFIENLLPSDGDWIGLR 78
Db 806 YKDYQYFSEKETMDNARAFKRNFGDLVSIQSESEKFLWKYI-NKNGGQSPYFIMGL 864
QY 79 RREEKOSNSTACODLYAWTDGSIQFRNMYVDPSGCS--EVCVMYTHQPSAPAGIGPY 136
Db 865 ISMDKK-----FIWMDGSKVDVFAWATGPNFANEDENCVTWY-----TNSGF--- 907
QY 137 MFQWNDRCNMKNFICK---YSDEKPAVPSREAEGEETELTPVLPEBTOE----- 185
Db 908 ---WINDINGCPYNNFICQRHNSINATAMP-----TPTTPPGCKEGWHLKYNK 953

QY 186 -----EDAKTKFKESREAAALNL 202
Db 954 CPKIFGFANEERKKSQWQARQACKGL 978

RESULT 3
S52781
neurocan - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 04-Feb-2000
C;Accession: S52781
R;Rauch, U.; Foreberg, N.; Kulbe, G.; Arnold-Ammer, I.; Paessler, R.
submitted to the EMBL Data Library, February 1995
A;Description: Amino acid sequence of mouse neurocan and brevicin and their different
A;Reference number: S52781
A;Accession: S52781
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1268 <RAU>
A;Cross-references: EMBL:X84727; NID:G758629; PIDN:CAA59216.1; PID:G758630
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
F;176-253/Domain: link protein repeat homology <LNK1>
F;274-355/Domain: link protein repeat homology <LNK2>
F;964-995/Domain: EGF homology <EGF>
F;1040-1160/Domain: C-type lectin homology <LCH>
F;1167-1223/Domain: complement factor H repeat homology <FHD>

Query Match 16.0%; Score 177; DB 2; Length 1268;
Best Local Similarity 31.5%; Pred. No. 2.1e-07;
Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;
QY 15 QRPCKYVIYFHTSRRLNFEAEKACRRDGGQVLVSIESEDEQKLIKFIENLLPSDGDW 74
Db 1048 QGHCYR--YF--AHRRWDAERDCRRRAGHLTSVHSPEEHKFINSF-----GHNSW 1096

QY 75 IGLRREKQSNSTACQDLYAWTDGSIQFRNMYVDPS---CGSEVCVMYTHQPSAPAG 131
Db 1097 IGLNDRTVERD-----FQWTDNTGLQYENREKQPNFPAGGEDCVVVAHESG--- 1145

QY 132 IGGPYWFQWDDRCNMKNFICK 154
Db 1146 -----RWNDVPCCYNLPYVCK 1161

RESULT 4
T14274
versican precursor, splice form V2 - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000
C;Accession: T14274
R;Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
J. Biol. Chem. 273, 15758-15764, 1998
A;Title: Versican V2 is a major extracellular matrix component of the mature bovine bra
A;Reference number: Z17954; MUID:98288320; PMID:9624174
A;Accession: T14274
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1643 <SCH>
A;Cross-references: EMBL:AF060458; NID:G3253303; PID:G3253304; PIDN:AAC24360.1
A;Experimental source: brain
C;Keywords: glycoprotein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-1643/Product: versican, splice form V2 #status predicted <MAT>
F;57,331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (Asn) (covalent) #

Query Match 15.8%; Score 174.5; DB 2; Length 1643;
Best Local Similarity 25.5%; Pred. No. 4.7e-07;
Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;
QY 15 QRPCKYVIYFHTSRRLNFEAEKACRRDGGQVLVSIESEDEQKLIKFIENLLPSDGD 73
Db 1424 QCCYK--YF--AHRRTWDAARECRLLQAGHLTSILSHEEQMFVNRV-----GHDYQ 1471

Qy 74 WIGLRREEKQSNSTACODLYAWTDGSIQFRNYYVDEP-----SCGSEVCMVYHQPAP 129
 |||||
 Db 1472 WIGL-----NDKMFHDFRWTGSLQYENWRPNQDFFSTGDCVVIWHENG-- 1521
 |||||
 Qy 130 AGIGPYMFQWDDRCNMKNFICKYS-----DEKPAVPSREAGE----- 170
 |||||
 Db 1522 -----QWNVPCNYHLTYTCKGTVACGQPPVVENAKTFGKMPRYEINSLIRYHC 1572
 |||||
 Qy 171 -----ETELTT-----PVL-----PEETQEDAKTFKESREAALN 201
 |||||
 Db 1573 KDGFIQRHLPTIRCLGNRWAMPKITCLNPSAYQRTYSKKYFNSSAKDN 1623
 |||||

RESULT 5
 T42389
 N;Alternate precursor, splice form V0 - bovine
 N;Alternate names: chondroitin sulfate proteoglycan
 C;Species: Bos primigenius taurus (cattle)
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 05-May-2000
 C;Accession: T42389
 R;Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
 J. Biol. Chem. 273, 15758-15764, 1998
 A;Title: Versican V2 is a major extracellular matrix component of the mature bovine brain
 A;Reference number: 217954; MUID:98288320; PMID:9624174
 A;Accession: T42389
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-3381 <SCH>
 A;Cross-references: EMBL:AF060456; NID:G3253299; PID:G3253300; PIDN:AAC24358.1
 C;Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lectin
 C;Keywords: chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein
 F;1-20/Domain: signal sequence #status predicted <SIG>
 F;21-3381/Product: versican, splice form V0 #status predicted <MAT>
 F;57,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2356,26

Query Match 15.8%; Score 174.5; DB 2; Length 3381;
 Best Local Similarity 25.5%; Pred. No. 1.1e-06;
 Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;
 Qy 15 QRPCYKVIYFHDTSRRRLNFEAEACRRDGGQLVSISEDEOKLIEKFIENLLPSDGDGF- 73
 |||||
 Db 3162 QGOCYK--YF--AHRRTWDAERECRLOQAHLTSLSHQEOMFVNRV-----GHYQ 3209
 |||||
 Qy 74 WIGLRREEKQSNSTACODLYAWTDGSIQFRNYYVDEP-----SCGSEVCMVYHQPAP 129
 |||||
 Db 3210 WIGL-----NDKMFHDFRWTGSLQYENWRPNQDFFSTGDCVVIWHENG-- 3259
 |||||
 Qy 130 AGIGPYMFQWDDRCNMKNFICKYS-----DEKPAVPSREAGE----- 170
 |||||
 Db 3260 -----QWNVPCNYHLTYTCKGTVACGQPPVVENAKTFGKMPRYEINSLIRYHC 3310
 |||||
 Qy 171 -----ETELTT-----PVL-----PEETQEDAKTFKESREAALN 201
 |||||
 Db 3311 KDGFIQRHLPTIRCLGNRWAMPKITCLNPSAYQRTYSKKYFNSSAKDN 3361
 |||||

RESULT 6
 S28764
 N;Alternate precursor - rat
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
 C;Accession: S28764
 R;Rauch, U.; Karthikeyan, L.; Maurel, P.; Margolis, R.U.; Margolis, R.K.
 J. Biol. Chem. 267, 19536-19547, 1992
 A;Title: Cloning and primary structure of neurocan, a developmentally regulated, aggregated proteoglycan
 A;Reference number: S28764; MUID:92406907; PMID:1326557
 A;Accession: S28764
 A;Molecule type: mRNA
 A;Residues: 1-1257 <RAU>
 A;Cross-references: EMBL:M97161; NID:G205649; PIDN:AAC37679.1; PID:G205650
 C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EG
 C;Keywords: chondroitin sulfate proteoglycan; glycoprotein

F;1-22/Domain: signal sequence #status predicted <SIG>
 F;23-1257/Product: neurocan #status predicted <MAT>
 F;176-253/Domain: link protein repeat homology <LNK1>
 F;274-355/Domain: link protein repeat homology <LNK2>
 F;364-366/Region: cell attachment (R-G-D) motif
 F;953-984/Domain: EGF homology <EGF>
 F;1029-1149/Domain: C-type lectin homology <LCH>
 F;1156-1212/Domain: complement factor H repeat homology <FHD>
 F;121,339,737,967,1164/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;372,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
 F;944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

Query Match 15.7%; Score 174; DB 2; Length 1257;
 Best Local Similarity 30.8%; Pred. No. 3.8e-07;
 Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;

Qy 15 QRPCYKVIYFHDTSRRRLNFEAEACRRDGGQLVSISEDEOKLIEKFIENLLPSDGDGF 74
 |||||
 Db 1037 QGHCYR--YF--AHRRAWEDAECDRCRRAGHLTSLSHPEKHKINSF-----GHENSW 1085
 |||||
 Qy 75 IGLRRREEKQSNSTACODLYAWTDGSIQFRNYYVDEPS---CGSEVCMVYHQPAP 131
 |||||
 Db 1086 IGLNDRITVEND-----FQWTDNTGLQYENWRKQPDNFFAGGEDCVVMVAHENG--- 1134
 |||||
 Qy 132 IGGPYMFQWDDRCNMKNFICK 154
 |||||
 Db 1135 -----RWNVPCNYNLPVCK 1150
 |||||

RESULT 7
 A55535
 N;Alternate precursor - mouse
 N;Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan
 N;Contains: glial hyaluronate-binding protein
 C;Species: Mus musculus (house mouse)
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C;Accession: A55535
 R;Ito, K.; Shinomura, T.; Zako, M.; Ujita, M.; Kimata, K.
 J. Biol. Chem. 270, 958-965, 1995
 A;Title: Multiple forms of mouse PG-M, a large chondroitin sulfate proteoglycan generated
 A;Reference number: A55535; MUID:95122551; PMID:7822336
 A;Accession: A55535
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-2397 <RES>
 C;Cross-references: GB:D16263; NID:G862460; PIDN:BAA03796.1; PID:G862461
 C;Superfamily: versican; C-type lectin homology; complement factor H repeat homology; I
 F;1-20/Domain: signal sequence #status predicted <SIG>
 F;21-1654/Domain: versican #status predicted <MAT>
 F;167-244/Domain: link protein repeat homology <LNK1>
 F;265-346/Domain: link protein repeat homology <LNK2>
 F;2095-2126/Domain: EGF homology <EG1>
 F;2133-2164/Domain: EGF homology <EG2>
 F;2171-2291/Domain: C-type lectin homology <LCH>
 F;2298-2354/Domain: complement factor H repeat homology <FHD>

Query Match 15.7%; Score 174; DB 1; Length 2397;
 Best Local Similarity 28.5%; Pred. No. 8e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

Qy 15 QRPCYKVIYFHDTSRRRLNFEAEACRRDGGQLVSISEDEOKLIEKFIENLLPSDGDGF- 73
 |||||
 Db 2179 QGOCYK--YF--AHRRTWDAERECRLOQAHLTSLSHQEOMFVNRV-----GHYQ 2226
 |||||
 Qy 74 WIGLRREEKQSNSTACODLYAWTDGSIQFRNYYVDEP-----SCGSEVCMVYHQPAP 129
 |||||
 Db 2227 WIGL-----NDKMFHDFRWTGSLQYENWRPNQDFFSAGEDCVVIWHENG-- 2276
 |||||
 Qy 130 AGIGPYMFQWDDRCNMKNFICKYS-----DEKPAVPSREAGE 170
 |||||
 Db 2277 -----QWNVPCNYHLTYTCKGTVACGQPPVVENAKTFGK 2312
 |||||

F:265-346/Domain: link protein repeat homology <LNK2>
F:559-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>
F:2106-2137/Domain: EGF homology <EG1>
F:2144-2175/Domain: EGF homology <EG2>
F:2182-2302/Domain: C-type lectin homology <LCH>
F:2309-2365/Domain: complement factor H repeat homology <PHD>

Query Match 15.7%; Score 174; DB 1; Length 2409;
Best Local Similarity 28.5%; Pred. No. 8e-07;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8

QY 15 QRPCKVIYFHDTSRRLNFEAKACRRDGQLVSISEDEQKLIEKFENLLPSGDGF-- 73
Db ||| | : | : | : | : | : | : | : | : | : | : | : | :
DB 2190 QQCYK--YF---AHRRTWDAERCRLOGAHLTSLSHSEQPVMNRV-----GHDYQ 223

QY 74 WIGLRREEKQSNTACODLYAWTDGISIFRNMYVDPE----SCGSEVCVMYHQPAP 129
Db ||| | : | : | : | : | : | : | : | : | : | : | : | :
DB 2238 WIGL-----NDKMFERFRWDGSTLOYENRPNQDSFFSAGEDCVIIWHENG-- 228

QY 130 AGIGGPYMFOWDDRCNMKNFIKYIS-----DEKPVPSPRAEAGE 170
Db ||| | : | : | : | : | : | : | : | : | : | : | : | :
DB 2288 -----QWNDFPCNYHLTCTCKGTACGGPPVVVENAKTFGK 2323

RESULT 9
A47171
chondroitin sulfate proteoglycan PG-M core protein - chicken
C:Species: Gallus gallus (chicken)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A47171
C:Shinomura, T.; Nishida, Y.; Ito, K.; Kimata, K.
J. Biol. Chem. 268, 14461-14469, 1993
A>Title: cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan expressed

A:Reference number: A47171; UID:93300846; PMID:8314802
A:Accession: A47171
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-3562 <SHI>
A:Cross-references: GB:D13542; NID:g391643; PIDN:BAA02742.1; PID:g391644
A:Experimental source: stage 22-23 developing limb buds
A>Note: sequence extracted from NCBI backbone (NCBIN:134456, NCBIPI:134457)
C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type
F:166-243/Domain: link protein repeat homology <LNK1>
F:264-345/Domain: link protein repeat homology <LNK2>
F:3258-3289/Domain: EGF homology <EGF1>
F:3296-3327/Domain: EGF homology <EGF>
F:3324-3454/Domain: C-type lectin homology <LCH>
F:3461-3511/Domain: complement factor H repeat homology <PHD>

Query Match 15.5%; Score 171; DB 2; Length 3562;
Best Local Similarity 28.5%; Pred. No. 2.3e-06;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8

QY 15 QRPCKVIYFHDTSRRLNFEAKACRRDGQLVSISEDEQKLIEKFENLLPSGDGF-- 73
Db ||| | : | : | : | : | : | : | : | : | : | : | : | :
DB 3342 QQCYK--YF---AHRRTWDAERCRLOGAHLTSLSHSEQPVMNR-----GHDYQ 3389

QY 74 WIGLRREEKQSNTACODLYAWTDGISIFRNMYVDPE----SCGSEVCVMYHQPAP 129
Db ||| | : | : | : | : | : | : | : | : | : | : | : | :
DB 3390 WIGL-----NDKMFERFRWDGSTLOYENRPNQDSFFSAGEDCVIIWHENG-- 3439

QY 130 AGIGGPYMFOWDDRCNMKNFIKYIS-----DEKPVPSPRAEAGE 170
Db ||| | : | : | : | : | : | : | : | : | : | : | : | :
DB 3440 -----QWNDFPCNYHLTCTCKGTACGGPPVVVENAKTFGK 3475

RESULT 10
T42710
mannose receptor, macrophage - mouse
N:Alternate names: lambda lectin; phospholipase A2 receptor
C:Species: Mus musculus (house mouse)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text change 09-Jun-2000

C;Accession: T42710
J;Wu, K.; Yuan, J.; Lasky, L.A.
R. Biol. Chem. 271, 21323-21330, 1996
A;Title: Characterization of a novel member of the macrophage mannose receptor type C le
A;Reference number: Z22235; MUID:5635501; PMID:8702911
A;Accession: T42710
A;Status: preliminary; translated from GB/EMBL/DDBB
A;Molecule type: mRNA
A;Residues: 1-1479 <WU>
A;Cross-references: EMBL: U56734; NID:g1336073; PID:g1336074; PIDN:AAC52729.1
C;Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II re
C;Keywords: membrane protein; receptor
F;186-227/Domain: fibronectin type II repeat homology <2FR>

Query Match 15.4%; Score 170.5; DB 2; Length 1479;
Best Local Similarity 31.4%; Pred.No. 9.3e-07;
Matches 48; Conservative 22; Mismatches 48; Indels 35; Gaps 7;

QY 5 SGQPVCRGGTGPCVKVIYFDTSRRLLNFEEAKECRRDGGQLVSTSEDEQKLIKFTIE 64
||| ||| : | : : | : | : | : | : | : | :
Db 384 SWQPF-----QGHCYRL-----QAEEKSWQESKRACLURGGDLLSIHSMAELFFITKQIK 433
:
QY 65 NLFPSDGDFWIGLRREERKEQSNTACODLYAWTDGSGISQFRNWYVDPEPS---CGSEVCVV 121
:
Db 434 QEVE---ELWIGL-----NDLKLMQNFENWDGSLVSFTHHPPEPNFRDSLEDCVT 482
:
QY 122 MYHQSPAPAGTGGPYMFQWNDRCKMNKNPICK 154
:
Db 483 IW----GPEG-----RWINDSPCNOSLPISICK 504
: : : : : : : : : : : : : : : : : : : :

RESULT 11

A39808

proteoglycan core protein, cartilage - bovine (fragments)

N;Alternate names: aggrecan; aggregating cartilage proteoglycan

C;Species: Bos primigenius taurus (cattle)

C;Date: 20-Mar-1992 #sequence revision 23-Mar-1995 #text change 13-Aug-1999

C;Accession: A34234; A27752; A39808; A27751; E29164; B27751; C27751; E27751; F27751; F27752; F27753; F27754; F27755; F27756; F27757; F27758; F27759; F27760; F27761; F27762; F27763; F27764; F27765; F27766; F27767; F27768; F27769; F27770; F27771; F27772; F27773; F27774; F27775; F27776; F27777; F27778; F27779; F27780; F27781; F27782; F27783; F27784; F27785; F27786; F27787; F27788; F27789; F27790; F27791; F27792; F27793; F27794; F27795; F27796; F27797; F27798; F27799; F27800; F27801; F27802; F27803; F27804; F27805; F27806; F27807; F27808; F27809; F27810; F27811; F27812; F27813; F27814; F27815; F27816; F27817; F27818; F27819; F27820; F27821; F27822; F27823; F27824; F27825; F27826; F27827; F27828; F27829; F27830; F27831; F27832; F27833; F27834; F27835; F27836; F27837; F27838; F27839; F27840; F27841; F27842; F27843; F27844; F27845; F27846; F27847; F27848; F27849; F27850; F27851; F27852; F27853; F27854; F27855; F27856; F27857; F27858; F27859; F27860; F27861; F27862; F27863; F27864; F27865; F27866; F27867; F27868; F27869; F27870; F27871; F27872; F27873; F27874; F27875; F27876; F27877; F27878; F27879; F27880; F27881; F27882; F27883; F27884; F27885; F27886; F27887; F27888; F27889; F27890; F27891; F27892; F27893; F27894; F27895; F27896; F27897; F27898; F27899; F27900; F27901; F27902; F27903; F27904; F27905; F27906; F27907; F27908; F27909; F27910; F27911; F27912; F27913; F27914; F27915; F27916; F27917; F27918; F27919; F27920; F27921; F27922; F27923; F27924; F27925; F27926; F27927; F27928; F27929; F27930; F27931; F27932; F27933; F27934; F27935; F27936; F27937; F27938; F27939; F27940; F27941; F27942; F27943; F27944; F27945; F27946; F27947; F27948; F27949; F27950; F27951; F27952; F27953; F27954; F27955; F27956; F27957; F27958; F27959; F27960; F27961; F27962; F27963; F27964; F27965; F27966; F27967; F27968; F27969; F27970; F27971; F27972; F27973; F27974; F27975; F27976; F27977; F27978; F27979; F27980; F27981; F27982; F27983; F27984; F27985; F27986; F27987; F27988; F27989; F27990; F27991; F27992; F27993; F27994; F27995; F27996; F27997; F27998; F27999; F28000; F28001; F28002; F28003; F28004; F28005; F28006; F28007; F28008; F28009; F28010; F28011; F28012; F28013; F28014; F28015; F28016; F28017; F28018; F28019; F28020; F28021; F28022; F28023; F28024; F28025; F28026; F28027; F28028; F28029; F28030; F28031; F28032; F28033; F28034; F28035; F28036; F28037; F28038; F28039; F28040; F28041; F28042; F28043; F28044; F28045; F28046; F28047; F28048; F28049; F28050; F28051; F28052; F28053; F28054; F28055; F28056; F28057; F28058; F28059; F28060; F28061; F28062; F28063; F28064; F28065; F28066; F28067; F28068; F28069; F28070; F28071; F28072; F28073; F28074; F28075; F28076; F28077; F28078; F28079; F28080; F28081; F28082; F28083; F28084; F28085; F28086; F28087; F28088; F28089; F28090; F28091; F28092; F28093; F28094; F28095; F28096; F28097; F28098; F28099; F28100; F28101; F28102; F28103; F28104; F28105; F28106; F28107; F28108; F28109; F28110; F28111; F28112; F28113; F28114; F28115; F28116; F28117; F28118; F28119; F28120; F28121; F28122; F28123; F28124; F28125; F28126; F28127; F28128; F28129; F28130; F28131; F28132; F28133; F28134; F28135; F28136; F28137; F28138; F28139; F28140; F28141; F28142; F28143; F28144; F28145; F28146; F28147; F28148; F28149; F28150; F28151; F28152; F28153; F28154; F28155; F28156; F28157; F28158; F28159; F28160; F28161; F28162; F28163; F28164; F28165; F28166; F28167; F28168; F28169; F28170; F28171; F28172; F28173; F28174; F28175; F28176; F28177; F28178; F28179; F28180; F28181; F28182; F28183; F28184; F28185; F28186; F28187; F28188; F28189; F28190; F28191; F28192; F28193; F28194; F28195; F28196; F28197; F28198; F28199; F28200; F28201; F28202; F28203; F28204; F28205; F28206; F28207; F28208; F28209; F28210; F28211; F28212; F28213; F28214; F28215; F28216; F28217; F28218; F28219; F28220; F28221; F28222; F28223; F28224; F28225; F28226; F28227; F28228; F28229; F28230; F28231; F28232; F28233; F28234; F28235; F28236; F28237; F28238; F28239; F28240; F28241; F28242; F28243; F28244; F28245; F28246; F28247; F28248; F28249; F28250; F28251; F28252; F28253; F28254; F28255; F28256; F28257; F28258; F28259; F28260; F28261; F28262; F28263; F28264; F28265; F28266; F28267; F28268; F28269; F28270; F28271; F28272; F28273; F28274; F28275; F28276; F28277; F28278; F28279; F28280; F28281; F28282; F28283; F28284; F28285; F28286; F28287; F28288; F28289; F28290; F28291; F28292; F28293; F28294; F28295; F28296; F28297; F28298; F28299; F28300; F28301; F28302; F28303; F28304; F28305; F28306; F28307; F28308; F28309; F2

A:Reference number: A91327; MUID:85027710; PMID:6489519
A:Accession: E29164
A:Molecule type: protein
A:Residues: 1230-1249 <PE2>
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
C:Keywords: glycoprotein
F:8-28/Domain: link protein repeat homology (fragment) <LNK1>
F:29-58/Domain: link protein repeat homology (fragment) <LNK2>
F:80-146/Domain: link protein repeat homology (fragments) <LNK3>
F:167-248/Domain: link protein repeat homology (fragments) <LNK4>
F:1130-1250/Domain: C-type lectin homology <LCH>
F:1257-1313/Domain: complement factor H repeat homology <PHD>

| | | | | |
|------------------------------|---------------|---------------------------|-----------------------|----------------------------|
| Query Match | 14.3% | Score 158.5; | DB 2; | Length 1340; |
| Best Local Similarity | 26.8%; | Pred. No. 9.ee-06; | | |
| Matches | 48; | Conservative 29; | Mismatches 59; | Indels 43; Gaps 10; |

| | | | | |
|-----------|-------------|---|--|-----------------|
| Qy | 7 | QPVCRGG--- | TQRPCYKVIVFHTDTSRLRLFPEAKBACRRGGQLGVLSIESDEDEKLIETKFI | 63 |
| | | | : : : : : : : : | : |
| Db | 1127 | QKLCEGWTKFOQHCVR--HFPP- | RATWVDRESOCRRQQSHLSSIVTPEEQ---- | EPV 1177 |
| | | | : : : : : : : : | : |
| Qy | 64 | ENLLPSDGDP-WTLGRRLREEKOSNSTACQDLVAWTDGSIQSFRNMYWVDSF---- | SCGSEV 118 | |
| | | | : : : : : : : : | : |
| Db | 1178 | NN---NAQDYQWIGL-----NDKTIEGDPRMSDGHSLQFNWRPNQDNFFATGEDC | 1226 | |
| | | | : : : : : : : : | : |
| Qy | 119 | CVVNYHQSPAPAGIGPYGFYNWDRCNNKNPICKYS-----DEKPAPVSRAEGETE | 173 | |
| | | | : : : : : : : : | : |
| Db | 1227 | VNMIMHEKG-----ENNWDVCNTQLPFTCKGTGKAVAGEPVVSHARIFGOKKD | 1274 | |
| | | | : : : : : : : : | : |

```

RESULT 12
T42630
aggrecan - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T42630
R:Hering, T.M.; Kollar, J.; Huynh, T.D.
submitted to the EMBL Data Library, September 1996
A:Description: Complete coding sequence of bovine aggrecan: comparative structural ana
A:Reference number: Z22182

```

A;Accession: I42650
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1-2337 CHER>
A;Cross-references: EMBL:U76615; NID:g1730259; PID:g1730260; PIDN:AAB38524.1
A;Experimental source: articular chondrocytes
C;Superfamily: agragan; C-type lectin homology; complement factor H repeat homology; 1
C;Keywords: cartilage; chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein

| Query Match | 14.3% | Score 158.5 | DB 2 | Length 2327 |
|-----------------------|-------|--|--|-------------|
| Best Local Similarity | 26.8% | Pred. No. 1.8e-05 | | |
| Matches | 48 | Conservative 29 | Mismatches 59 | Indels 43 |
| | | | | Gaps 10 |
| QY | 7 | QPVCRGG--- | TQPCVKVIYHFHDSRLNFEAKACRRDGGQLVYSISEDEOKLIERFI | 63 |
| | | | | |
| Db | 2114 | QKLREGWTKFOHCYR---HFPPD--- | RATWVDAESQCRKQSHLSIVTPEEQ---- | BFV 2164 |
| | | | | |
| QY | 64 | ENLLPSDGF-WIGLRREEKQSNSTACQDLYAWTDGSI | SQFRNMYVDEP----- | SCGSEV 118 |
| | | | | |
| Db | 2165 | NN---NAQDYQWIGL----- | NDKTIEGDFRWSGDGHSLOFENWRPNQDPNFPATGEDC | 2213 |
| | | | | |
| QY | 119 | CVVNTHQSPAGIGGYPWFQWDDRCNNKONFICKYS----- | DEKPAVPSREAGEETS | 173 |
| | | | | |
| Db | 2214 | VVMIVHEKRG----- | ENWDVPCYNQLPFTCKGTGTVACGGPPVVEHARIFQOKKD | 2261 |
| | | | | |

RESULT 13
A39086
aggrecan precursor, cartilage long splice form [validated] - human
N;Alternate names: chondroitin sulfate proteoglycan 1; large aggregating proteoglycan;
N;Contains: aggrecan cartilage short splice form
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence revision 01-Dec-2000 #text change 08-Dec-2000

F;1-21/Domain: signal sequence #status predicted <SIG>
F;12-138/Domain: C-type lectin homology <LCH>
F;22-612/Product: P-selectin #status predicted <MAT>
F;143-174/Domain: EGF homology <EGF>
F;180-238/Domain: complement factor H repeat homology <FHI>
F;243-300/Domain: complement factor H repeat homology <FH2>
F;305-363/Domain: complement factor H repeat homology <FH3>
F;368-426/Domain: complement factor H repeat homology <FH4>
F;431-489/Domain: complement factor H repeat homology <FH5>
F;494-548/Domain: complement factor H repeat homology <FH6>
F;25,391,528/Binding site: carbohydrate (Asn) #status predicted

Query Match 14.0%; Score 154.5; DB 2; Length 612;
Best Local Similarity 27.9%; Pred. No. 8,7e-06;
Matches 38; Conservative 28; Mismatches 45; Indels 25; Gaps 5;

Qy 23 YFHTDTRLNPFEEAKACRRGGOLVSIESEDEOKLIEKFIENLLPSGDFWIGLRRREE 82
Db YN::::::|||||:::|||||:::|||||:::|||||:::|||||::: 76
23 YNASSLSMTYDEASAYCORDYTHLVAIONKEE---INLYNSNLKHSPSYVWIGIRK--- 76

Qy 83 KQSNSTACQDLXAWTDGS---ISQFRNMYVDEPS--CGSEVGVVMVHQPAPAGIGPGYM 137
Db :::::|||||:::|||||:::|||||:::|||||:::|||||::: 124
77 -----VNNVWVGTKPLTEAQNAWGPENKRNKORNCDEVIYIQRKTDSGM----- 124

Qy 138 FQWDDRCNNKKNFIC 153
Db :::|||||::: 125 --WNDERCNKKLALC 138

RESULT 15
A28452
proteoglycan core protein precursor, cartilage - rat
N;Alternate names: aggrecan
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1989 #sequence revision 30-Jun-1989 #text_change 21-Jan-2000
C;Accession: A92623; A23835; A28453; A28095; A28452
R;Doegge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.
J. Biol. Chem. 262, 17757-17767, 1987
A;Title: Complete primary structure of the rat cartilage proteoglycan core protein deduced from complementary DNA sequence
A;Reference number: A92623; MUID:88087070; PMID:3693370
A;Accession: A28452
A;Molecule type: mRNA
A;Residues: 1-2124 <DOE>
R;Doegge, K.; Sasaki, M.; Horigan, E.; Hassell, J.R.; Yamada, Y.
J. Biol. Chem. 263, 10040a, 1988
A;Reference number: A30069
A;Contents: annotation; revision to residue 698
R;Doegge, K.; Fernandez, P.; Hassell, J.R.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 261, 8108-8111, 1986
A;Title: Partial cDNA sequence encoding a globular domain at the C terminus of the rat cartilage proteoglycan core protein deduced from complementary DNA sequence
A;Reference number: A23835; MUID:86250698; PMID:2424893
A;Accession: A23835
A;Molecule type: mRNA
A;Residues: 1856-2124 <DO2>
A;Cross-references: GB:M13518; NID:g206104; PIDN:AAA41836.1; PID:g206105
R;Neame, P.J.; Christner, J.E.; Baker, J.R.
J. Biol. Chem. 262, 17768-17778, 1987
A;Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan amino-terminal domain
A;Reference number: A28453; MUID:88087071; PMID:3693371
A;Accession: A28453
A;Molecule type: protein
A;Residues: 20-37, 'W', 39-60, 'E', 62-64, 'X', 66-69, 70-83, 84, 89-148, 'L', 150-238, 'S', 240, 'A', 'C';Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; E;Keywords: glycoprotein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-2124/Product: proteoglycan core protein #status predicted <MAT>
F;44-135/Domain: immunoglobulin homology <IMM>
F;170-247/Domain: link protein repeat homology <LNK1>
F;268-349/Domain: link protein repeat homology <LNK2>
F;504-581/Domain: link protein repeat homology <LNK3>
F;602-683/Domain: link protein repeat homology <LNK4>
F;1914-2034/Domain: C-type lectin homology <LCH>
F;2041-2097/Domain: complement factor H repeat homology <FHD>

```

F:126,239,333,387,611,667,1042/Binding site: carbohydrate (Aen) (covalent) #status pr
Query Match      13.9%; Score 153.5; DB 2; Length 2124;
Best Local Similarity 26.3%; Pred. No. 4.5e-05;
Matches 47; Conservative 27; Mismatches 62; Indels 43; Gaps 9;

Qy 7 QPVCRRG---TORPCYKVIYFHDTSRRLLNFEEAKEACRRDGGGOLVSIESDEQKLIIEKFI 63
Db 1911 QEQCEEKWTKEFGHCYR--HFPD---RETWVDAERRCRREQQSHLSIVTPEEQEFVNKNA 1965

Qy 64 ENLLPSDGF-WIGLRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEP----SCGSEV 118
Db 1966 Q-----DYQWIGL-----NDRITIEGDFRWSGDHSLQFEKWRPNQDNFFATGEDC 2010

Qy 119 CVVMYHOPSAFAGIGPYMFQWNDORCMNNFNICKYS-----DEKPAVPSREAEGETE 173
Db 2011 VVMIWHERG-----EWNQVPCNYQLPFTCKKGTGTVACGEPFAVEHARTLGOKKD 2058

Search completed: December 22, 2003, 16:14:41
Job time : 10.9179 secs

```

Search completed: December 22, 2003, 16:14:41
Job time : 10.9179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:07:10 ; Search time 19.4937 Seconds
(without alignments)
1954.412 Million cell updates/sec

Title: US-09-887-855-2_COPY_24_227

Perfect score: 1106

Sequence: 1 GRLLSGQPVCRGGTQPCYK.....EEDAKTKFKESREAAALNLAY 204

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 1106 | 100.0 | 206 | 9 | US-09-887-855-5 |
| 2 | 1106 | 100.0 | 374 | 9 | US-09-887-855-2 |
| 3 | 1106 | 100.0 | 374 | 12 | US-10-094-749-2090 |
| 4 | 1106 | 100.0 | 374 | 15 | US-10-149-819-15 |
| 5 | 1102 | 99.6 | 374 | 12 | US-10-094-749-2142 |
| 6 | 1092 | 98.7 | 382 | 10 | US-09-909-320-137 |
| 7 | 1092 | 98.7 | 382 | 10 | US-09-909-088B-137 |
| 8 | 1092 | 98.7 | 382 | 10 | US-09-905-291A-137 |
| 9 | 1092 | 98.7 | 382 | 10 | US-09-902-853-137 |
| 10 | 1092 | 98.7 | 382 | 10 | US-09-907-824-137 |
| 11 | 1092 | 98.7 | 382 | 10 | US-09-907-841-137 |
| 12 | 1092 | 98.7 | 382 | 11 | US-09-904-011-137 |
| 13 | 1092 | 98.7 | 382 | 11 | US-09-906-742-137 |
| 14 | 1092 | 98.7 | 382 | 11 | US-09-906-838-137 |
| 15 | 1092 | 98.7 | 382 | 11 | US-09-907-613-137 |

| | | | | | | |
|----|------|------|-----|----|--------------------|-------------------|
| 16 | 1092 | 98.7 | 382 | 11 | US-09-907-942-137 | Sequence 137, App |
| 17 | 1092 | 98.7 | 382 | 11 | US-09-904-859-137 | Sequence 137, App |
| 18 | 1092 | 98.7 | 382 | 11 | US-09-909-204-137 | Sequence 137, App |
| 19 | 1092 | 98.7 | 382 | 11 | US-09-904-820-137 | Sequence 137, App |
| 20 | 1092 | 98.7 | 382 | 11 | US-09-904-786-137 | Sequence 137, App |
| 21 | 1092 | 98.7 | 382 | 11 | US-09-906-646-137 | Sequence 137, App |
| 22 | 1092 | 98.7 | 382 | 11 | US-09-906-700-137 | Sequence 137, App |
| 23 | 1092 | 98.7 | 382 | 11 | US-09-903-786-137 | Sequence 137, App |
| 24 | 1092 | 98.7 | 382 | 11 | US-09-902-903-137 | Sequence 137, App |
| 25 | 1092 | 98.7 | 382 | 11 | US-09-903-749A-137 | Sequence 137, App |
| 26 | 1092 | 98.7 | 382 | 11 | US-09-904-119-137 | Sequence 137, App |
| 27 | 1092 | 98.7 | 382 | 11 | US-09-904-956-137 | Sequence 137, App |
| 28 | 1092 | 98.7 | 382 | 11 | US-09-902-736-137 | Sequence 137, App |
| 29 | 1092 | 98.7 | 382 | 11 | US-09-907-794-137 | Sequence 137, App |
| 30 | 1092 | 98.7 | 382 | 11 | US-09-903-943-137 | Sequence 137, App |
| 31 | 1092 | 98.7 | 382 | 11 | US-09-904-462-137 | Sequence 137, App |
| 32 | 1092 | 98.7 | 382 | 11 | US-09-907-525-137 | Sequence 137, App |
| 33 | 1092 | 98.7 | 382 | 11 | US-09-902-692-137 | Sequence 137, App |
| 34 | 1092 | 98.7 | 382 | 11 | US-09-903-520-137 | Sequence 137, App |
| 35 | 1092 | 98.7 | 382 | 11 | US-09-905-056-137 | Sequence 137, App |
| 36 | 1092 | 98.7 | 382 | 11 | US-09-909-064-137 | Sequence 137, App |
| 37 | 1092 | 98.7 | 382 | 11 | US-09-904-553-137 | Sequence 137, App |
| 38 | 1092 | 98.7 | 382 | 11 | US-09-905-381-137 | Sequence 137, App |
| 39 | 1092 | 98.7 | 382 | 11 | US-09-905-088-137 | Sequence 137, App |
| 40 | 1092 | 98.7 | 382 | 11 | US-09-907-575-137 | Sequence 137, App |
| 41 | 1092 | 98.7 | 382 | 11 | US-09-905-075-137 | Sequence 137, App |
| 42 | 1092 | 98.7 | 382 | 11 | US-09-902-759-137 | Sequence 137, App |
| 43 | 1092 | 98.7 | 382 | 11 | US-09-902-634-137 | Sequence 137, App |
| 44 | 1092 | 98.7 | 382 | 11 | US-09-902-713-137 | Sequence 137, App |
| 45 | 1092 | 98.7 | 382 | 11 | US-09-907-979-137 | Sequence 137, App |

ALIGNMENTS

RESULT 1
US-09-887-855-5
; Sequence 5, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-887-855-5

Query Match 100.0%; Score 1106; DB 9; Length 206;
Best Local Similarity 100.0%; Pred. No. 6.7e-104;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | | |
|----|-----|--|--------|-----|
| Qy | 1 | GRLLSGQPVCRGGTQPCYKVIYFHDTSRLNFEAKACRRDGGQLVSI | DEQKLI | 60 |
| Db | 3 | GRLLSGQPVCRGGTQPCYKVIYFHDTSRLNFEAKACRRDGGQLVSI | DEQKLI | 62 |
| Qy | 61 | KFTENLLPSDGDWIGLRRREKXSNSTACODLYAWTDGSIQPRN | WYDPSG | 120 |
| Db | 63 | KFTENLLPSDGDWIGLRRREKXSNSTACODLYAWTDGSIQPRN | WYDPSG | 122 |
| Qy | 121 | VMYHQSPAPAGIGGPFYFQWDDRCNNKNFICKYSDKPAVPSRE | AGTETL | 180 |
| Db | 123 | VMYHQSPAPAGIGGPFYFQWDDRCNNKNFICKYSDKPAVPSRE | AGTETL | 182 |
| Qy | 181 | EETQEDAKTKFKESREAAALNLAY | 204 | |
| Db | 183 | EETQEDAKTKFKESREAAALNLAY | 206 | |

```

; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2090
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2090

Query Match      100.0%; Score 1106; DB 12; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.5e-103;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy      1  GRLLSGQPVCRGGTQRPCYKVIYPHDTSRRLNFPFEAKACRRDGGQLVSISEDEQKLI 60
Db      24  GRLLSGQPVCRGGTQRPCYKVIYPHDTSRRLNFPFEAKACRRDGGQLVSISEDEQKLI 83

Qy      61  KFTENLLPSDGDGFWIGURRREEKQSNSTACODLYAWTDGSIQFRNMVYVDEPSCGSEVCV 120
Db      84  KFTENLLPSDGDGFWIGURRREEKQSNSTACODLYAWTDGSIQFRNMVYVDEPSCGSEVCV 143

Qy      121  VMYHQPSAPAGIGGPFYMFQWDDRCNKNKNPICKYSDKPAVPSPREAGEGETELTTPVLP 180
Db      144  VMYHQPSAPAGIGGPFYMFQWDDRCNKNKNPICKYSDKPAVPSPREAGEGETELTTPVLP 203

Qy      181  EETQEDAKKTFKESREAAALNAY 204
Db      204  EETQEDAKKTFKESREAAALNAY 227

```

```

RESULT 4
US-10-149-819-15
; Sequence 15, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUB, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: SHAH, Purvi
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: EP-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 374
; TYPE: PRP
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 3143411CD1
US-10-149-819-15

```

Qy 121 VMYHPSAPAGIGGYPYFQWDDRCNMKNFICKYSDKPAVPSREAGEETELTPVLP 180
Db 144 VMYHPSAPAGIGGYPYFQWDDRCNMKNFICKYSDKPAVPSREAGEETELTPVLP 203
Qy 181 EETQEDAKTKFKESREAAALNAY 204
Db 204 EETQEDAKTKFKESREAAALNAY 227

RESULT 5

US-10-094-749-2142
; Sequence 2142, Application US/10094749
; Publication No. US20030219741A1

; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: HIO, YURI

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHICO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOTYUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA

; FILE REFERENCE: 084335/0160

; CURRENT APPLICATION NUMBER: US/10/094,749

; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 60/350,435

; PRIOR FILING DATE: 2002-01-24

; PRIOR APPLICATION NUMBER: JP 2001-328381

; PRIOR FILING DATE: 2001-09-14

; NUMBER OF SEQ ID NOS: 3381

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2142

; LENGTH: 374

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-094-749-2142

Query Match 99.6%; Score 1102; DB 12; Length 374;
Best Local Similarity 99.5%; Pred. No. 3,8e-103;
Matches 203; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GRLLSGQVCRGGTQRCYKVIYPHDTSRRLNFEFEAKACRRDGGQLVSEDEQKLE 60
Db 24 GRLLSGQVCRGGTQRCYKVIYPHDTSRRLNFEFEAKACRRDGGQLVSEDEQKLE 83

Qy 61 KFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQPRNMYVDEPSCGSEVCV 120
Db 84 KFIENLLPSDGFWMGLRRREKQSNSTACQDLYAWTDGSIQPRNMYVDEPSCGSEVCV 143

Qy 121 VMYHPSAPAGIGGYPYFQWDDRCNMKNFICKYSDKPAVPSREAGEETELTPVLP 180
Db 144 VMYHPSAPAGIGGYPYFQWDDRCNMKNFICKYSDKPAVPSREAGEETELTPVLP 203

Qy 181 EETQEDAKTKFKESREAAALNAY 204
Db 204 EETQEDAKTKFKESREAAALNAY 227

RESULT 6

US-09-909-320-137

; Sequence 137, Application US/09909320

; Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas P.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-320-137


```
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905/291A
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-291A-137

Query Match          98.7%; Score 1092; DB 10; Length 382;
Best Local Similarity 96.2%; Pred. No. 4e-102;
Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 GRLLS-----GQPCVCGGTGTCVKYVYHDTSRRLNFEAKEACRDGGQLVSIES 52
Db 24 GRLLSASDLRGQPCVCGGTGTCVKYVYHDTSRRLNFEAKEACRDGGQLVSIES 83
QY 53 EDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTSGISQFRNMYVDEP 112
Db 84 EDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTSGISQFRNMYVDEP 143
QY 113 SCGSEVCVMYHQPAPAGIGGPMYFQWDDRCNMKNFICKYSDEKPAVPSREAGEET 172
Db 144 SCGSEVCVMYHQPAPAGIGGPMYFQWDDRCNMKNFICKYSDEKPAVPSREAGEET 203
QY 173 ELTPVLPETQEDAKTKFKESREAAALNLAY 204
Db 204 ELTPVLPETQEDAKTKFKESREAAALNLAY 235

RESULT 9
US-09-902-853-137
; Sequence 137, Application US/09902853
; Publication No. US20020192659A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Goddard, A.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavir, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas P.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,853
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US/09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-902-853-137

Query Match          98.7%; Score 1092; DB 10; Length 382;
```

```

Best Local Similarity 96.2%; Pred. No. 4e-102; Mismatches 0; Indels 8; Gaps 1;
Matches 204; Conservative 0;

Qy 1 GRLLS-----QPVCRGGTQPCVKIYFHDTSRRLNPFEEAKEACRRDGGOLVSIES 52
    |||||
Db 24 GRLLSASDLRLRGQPVCRGGTQPCVKIYFHDTSRRLNPFEEAKEACRRDGGOLVSIES 83
    |||||
Qy 53 EDEOKLIEKFIENLLPSDGFWICLRREEKQSNSTACQDLYAWTDGSIQFRNWWYDEP 112
    |||||
Db 84 EDEOKLIEKFIENLLPSDGFWICLRREEKQSNSTACQDLYAWTDGSIQFRNWWYDEP 143
    |||||
Qy 113 SCGSEVCVMYHQPAPAGIGGPMFQWDDRCNMKNFKICKYSDKPAVPSREAEGET 172
    |||||
Db 144 SCGSEVCVMYHQPAPAGIGGPMFQWDDRCNMKNFKICKYSDKPAVPSREAEGET 203
    |||||
Qy 173 ELTPVLPETQEDAKKTFKESREAAALNLAY 204
    |||||
Db 204 ELTPVLPETQEDAKKTFKESREAAALNLAY 235
    |||||

```

RESULT 10

```

US-09-907-824-137
; Sequence 137, Application US/09907824
; Publication No. US20020197671A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,824
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15

```

```

; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-907-824-137

Query Match 98.7%; Score 1092; DB 10; Length 382;
Best Local Similarity 96.2%; Pred. No. 4e-102; Indels 8; Gaps 1;
Matches 204; Conservative 0; Mismatches 0;

Qy 1 GRLLS-----QPVCRGGTQPCVKIYFHDTSRRLNPFEEAKEACRRDGGOLVSIES 52
    |||||
Db 24 GRLLSASDLRLRGQPVCRGGTQPCVKIYFHDTSRRLNPFEEAKEACRRDGGOLVSIES 83
    |||||
Qy 53 EDEOKLIEKFIENLLPSDGFWICLRREEKQSNSTACQDLYAWTDGSIQFRNWWYDEP 112
    |||||
Db 84 EDEOKLIEKFIENLLPSDGFWICLRREEKQSNSTACQDLYAWTDGSIQFRNWWYDEP 143
    |||||
Qy 113 SCGSEVCVMYHQPAPAGIGGPMFQWDDRCNMKNFKICKYSDKPAVPSREAEGET 172
    |||||
Db 144 SCGSEVCVMYHQPAPAGIGGPMFQWDDRCNMKNFKICKYSDKPAVPSREAEGET 203
    |||||
Qy 173 ELTPVLPETQEDAKKTFKESREAAALNLAY 204
    |||||
Db 204 ELTPVLPETQEDAKKTFKESREAAALNLAY 235
    |||||

```

RESULT 11

```

US-09-907-841-137
; Sequence 137, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.

```


; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,941
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-907-841-137

Query Match 98.7%; Score 1092; DB 10; Length 382;
Best Local Similarity 96.2%; Pred. No. 4e-102;
Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 GRLLS-----GQPVCRGQTPCYKVIYFHDTSRLNFEFAKEACRRDGGQVSVIES 52
Db 24 GRLLSADLDLRGQPVCRGQTPCYKVIYFHDTSRLNFEFAKEACRRDGGQVSVIES 83

QY 53 EDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEP 112
Db 84 EDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEP 143

QY 113 SCGSEVCVMYHQPSPAGIGGPTWFMQNDRCNMKNFNICKYSDEKPAVPSREAEGEET 172
Db 144 SCGSEVCVMYHQPSPAGIGGPTWFMQNDRCNMKNFNICKYSDEKPAVPSREAEGEET 203

QY 173 ELTPVLPPEETOEDAKTKFESREAAALNAY 204
Db 204 ELTPVLPPEETOEDAKTKFESREAAALNAY 235

RESULT 12
US-09-904-011-137
; Sequence 137, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Denoyere, Luc
; APPLICANT: Saton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-904-011-137

Query Match 98.7%; Score 1092; DB 11; Length 382;
Best Local Similarity 96.2%; Pred. No. 4e-102;
Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 GRLLS-----GQPVCRGQTPCYKVIYFHDTSRLNFEFAKEACRRDGGQVSVIES 52
Db 24 GRLLSADLDLRGQPVCRGQTPCYKVIYFHDTSRLNFEFAKEACRRDGGQVSVIES 83

QY 53 EDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEP 112
Db 84 EDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEP 143

APPLICANT: Filvaroff, Ellen
APPLICANT: Pong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kijavini, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US 09/907,613
CURRENT FILING DATE: 2001-07-17
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 137
LENGTH: 382
TYPE: PRT
ORGANISM: Homo Sapien
US-09-906-838-137

Query Match 98.7%; Score 1092; DB 11; Length 382;
Best Local Similarity 96.2%; Pred. No. 4e-102;
Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 1 GRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSIES 52
DB 24 GRLLSASDLRLGGQPCVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSIES 83
QY 53 EDEOKLIEKFLENLLPSDGFHWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEP 112
DB 84 EDEOKLIEKFLENLLPSDGFHWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEP 143
QY 113 SCGSEVCVMYHQPSAPAGIGPYFQWDDRCNMKNNFICKYSDKEKPAVPSREAGSET 172
DB 144 SCGSEVCVMYHQPSAPAGIGPYFQWDDRCNMKNNFICKYSDKEKPAVPSREAGSET 203
QY 173 ELTPVLPETQEDAKKTFKESREAAALNLAY 204
DB 204 ELTPVLPETQEDAKKTFKESREAAALNLAY 235

RESULT 15
US-09-907-613-137
Sequence 137, Application US/09907613
Publication No. US20030027145A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Deenoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone

Query Match 98.7%; Score 1092; DB 11; Length 382;
Best Local Similarity 96.2%; Pred. No. 4e-102;
Matches 204; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 1 GRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSIES 52
DB 24 GRLLSASDLRLGGQPCVCRGGTQPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSIES 83

| | | | | | |
|----|-----|---------------------|-----------------------------|------------------------|-----|
| Qy | 53 | EDEQKLEKFIENLLPSDGF | WIGLRREEKQSNSTACQDLYAWTDGSI | SQFRNYYVDEP | 112 |
| Db | 84 | EDEQKLEKFIENLLPSDGF | WIGLRREEKQSNSTACQDLYAWTDGSI | SQFRNYYVDEP | 143 |
| Qy | 113 | SCGSEVCVVMYHQP | SAPAGIGGPFYMFQWDDRCNMKN | NFICKYSDEKPAVPSREAGEET | 172 |
| Db | 144 | SCGSEVCVVMYHQP | SAPAGIGGPFYMFQWDDRCNMKN | NFICKYSDEKPAVPSREAGEET | 203 |
| Qy | 173 | ELTTPVLPEETQ | EEDAKKTFKESREAA | NLAY | 204 |
| Db | 204 | ELTTPVLPEETQ | EEDAKKTFKESREAA | NLAY | 235 |

Search completed: December 22, 2003, 16:16:47
Job time : 19.4937 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:05:09 ; Search time 6.56161 Seconds
(without alignments)
1476.391 Million cell updates/sec

Title: US-09-887-855-2_COPY_22_227

Perfect score: 1115
Sequence: 1 ATGRLSGQPVCRGQTRPC.....EEDAKTKFSREAAALNAY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 574 | 51.5 | 273 | 1 | Q9cxb0 mus musculus |
| 2 | 561.5 | 50.4 | 273 | 1 | CHOD_MOUSE |
| 3 | 182 | 16.3 | 1456 | 1 | MANR_HUMAN |
| 4 | 177 | 15.9 | 1268 | 1 | PGCN_MOUSE |
| 5 | 176 | 15.8 | 1321 | 1 | PGCN_HUMAN |
| 6 | 174.5 | 15.7 | 3381 | 1 | PGCV_BOVIN |
| 7 | 174 | 15.6 | 1257 | 1 | PGCV_RAT |
| 8 | 174 | 15.6 | 2738 | 1 | PGCV_RAT |
| 9 | 174 | 15.6 | 3358 | 1 | PGCV_MOUSE |
| 10 | 174 | 15.6 | 3396 | 1 | PGCV_HUMAN |
| 11 | 171 | 15.3 | 3562 | 1 | PGCV_CHICK |
| 12 | 165 | 14.8 | 643 | 1 | CD93_RAT |
| 13 | 158.5 | 14.2 | 2364 | 1 | PGCA_BOVIN |
| 14 | 158.5 | 14.2 | 2415 | 1 | PGCA_HUMAN |
| 15 | 155.5 | 13.9 | 2333 | 1 | PGCA_CANFA |
| 16 | 154.5 | 13.9 | 612 | 1 | LEM2_MOUSE |
| 17 | 153.5 | 13.8 | 644 | 1 | CD93_MOUSE |
| 18 | 153.5 | 13.8 | 2124 | 1 | PGCA_RAT |
| 19 | 152 | 13.6 | 912 | 1 | PGCB_BOVIN |
| 20 | 151.5 | 13.6 | 652 | 1 | CD93_HUMAN |
| 21 | 151 | 13.5 | 197 | 1 | CLPT_HUMAN |
| 22 | 149 | 13.4 | 321 | 1 | FCB2_HUMAN |
| 23 | 148.5 | 13.3 | 2132 | 1 | PGCA_MOUSE |
| 24 | 147 | 13.2 | 883 | 1 | PGCB_MOUSE |
| 25 | 146 | 13.1 | 2109 | 1 | PGCA_CHICK |
| 26 | 145 | 13.0 | 158 | 1 | LECG_TRIST |
| 27 | 145 | 13.0 | 883 | 1 | PGCB_RAT |
| 28 | 144.5 | 13.0 | 173 | 1 | LEC2_MEGRO |
| 29 | 144.5 | 13.0 | 372 | 1 | LEM1_RAT |
| 30 | 143.5 | 12.9 | 372 | 1 | LEM1_MOUSE |
| 31 | 141.5 | 12.7 | 331 | 1 | FCE2_MOUSE |
| 32 | 141.5 | 12.7 | 549 | 1 | LEM2_RAT |
| 33 | 140.5 | 12.6 | 162 | 1 | LEC3_MEGRO |
| | | | | | F07439 megabalanus |

| | | | | | | |
|----|-------|------|-----|---|------------|---------------------|
| 34 | 140 | 12.6 | 370 | 1 | LEM1_BOVIN | P98131 bos taurus |
| 35 | 139 | 12.5 | 248 | 1 | PSPA_HUMAN | P07714 homo sapien |
| 36 | 139 | 12.5 | 283 | 1 | LECA_SARPE | P05047 sarcophaga |
| 37 | 138.5 | 12.4 | 152 | 1 | IXA_TRIFL | P23806 trimeresuru |
| 38 | 138.5 | 12.4 | 202 | 1 | TETN_MOUSE | P43025 mus musculus |
| 39 | 137.5 | 12.3 | 372 | 1 | LEM1_MACMU | Q95198 macaca mula |
| 40 | 137.5 | 12.3 | 372 | 1 | LEM1_PAPHA | Q28768 papio hamad |
| 41 | 136 | 12.2 | 175 | 1 | LITH_BOVIN | P23132 bos taurus |
| 42 | 135 | 12.1 | 165 | 1 | LIT1_MOUSE | P43137 mus musculus |
| 43 | 134.5 | 12.1 | 132 | 1 | ACAL_ANGAN | P83300 anser anser |
| 44 | 134.5 | 12.1 | 372 | 1 | LEM1_PONPY | Q95235 pongo pygma |
| 45 | 134.5 | 12.1 | 485 | 1 | LEM2_BOVIN | P98107 bos taurus |

ALIGNMENTS

RESULT 1
CHOD_MOUSE
ID CHOD_MOUSE STANDARD; PRT; 273 AA.
AC Q9CXB0; 08V131;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chondrolectin precursor (Transmembrane protein MT75).
GN CHODL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Weng L., Smits P., Hubner R., Wouters J., Merregaert J.;
RT "Mt75, a low expressed c-type lectin gene involving in
RT chondrogenesis";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Carninci P., de Bonaldo M.F.,
RA Gwinnich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momabaets P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontesaki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF311699; AAL50354.1; -.

DR EMBL; AK014255; BAB29226.1; -.
 DR HSSP; P22897; IEGG.
 DR MGD; MGI:2179069; Chodl.
 DR InterPro; IPR001304; Lectin_C.
 DR Pfam; PF00059; lectin_c; 1.
 DR SMART; SM00034; CLECT; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; FALSE_NEG.
 DR PROSITE; PS50041; C-TYPE LECTIN 2; 1.
 KW Lectin; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 273 CHONDROLECTIN.
 FT DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 217 237 POTENTIAL.
 FT DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 35 179 C-TYPE LECTIN.
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 24 24 V -> W (IN REF. 2).
 FT CONFLICT 179 179 T -> K (IN REF. 2).
 SQ SEQUENCE 273 AA; 30303 MW; E052D933F244F4C7 CRC64;
 Query Match 51.5%; Score 574; DB 1; Length 273;
 Best Local Similarity 57.4%; Pred. No. 4.3e-44;
 Matches 109; Conservative 26; Mismatches 41; Indels 14; Gaps 5;
 QY 4 RLISGQVRCGRTQPCYKVIYHDTSRRLNFEAKACRRDQGLVSISEDEKLIEX 63
 Db 23 RVVSGQKVCADVKHPCYKMAYPHELSSRVFQEARLACRSEGGLVLSLENAEQKLI 82
 QY 64 FIENLLP-----SDGDFWIGLRREKQNSTACODLYANTDGSISQFRWYVDPSG 118
 Db 83 MLQNLTKPTGIGSDGDFWIGLRSRGQGT-SGACPDLYQWSDGSSQFRWYVDPSG 141
 QY 119 EVCVMVYHQPAPAGIGGYPFMFQWDDRCNNKNFICKYSDE-KPAVPSREASGEETELT 177
 Db 142 EKVVMVYHQTANPGLGGPLYQWDDRCNNKNYICTYPEIHTPEPA-----EXPLYLT 196
 QY 178 TPVLPEPTEQ 187
 Db 197 NQ--PEETHE 204
 RESULT 2
 CHOD_HUMAN STANDARD; PRT; 273 AA.
 AC Q9H9P2; Q9HCY3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Chondrolectin precursor (Transmembrane protein MT75) (PRED12 protein).
 GN CHODL OR C21ORF68.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RX MEDLINE=22074930; PubMed=12079284;
 RA Weng L., Smits P., Wauters J., Merregaert J.;
 RT "Molecular cloning and characterization of human chondrolectin, a novel type I transmembrane protein homologous to C-type lectins."; Genomics 80:62-70(2002).
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y., Soeda E., Onki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Follert A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E., Minoshima S., Shimizu N., Nordtsiek G., Hornischer K., Brandt P.,

RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H., Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehman H., Reinhardt R., Yaepo M.-D.;
 RT "The DNA sequence of human chromosome 21."; Nature 405:311-319(2000).
 RN [3]
 RN SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Haiech F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Sapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RN SEQUENCE OF 27-273 FROM N.A.
 RA Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Wagaetsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Ninomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project."; Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN TISSUE SPECIFICITY.
 RX MEDLINE=215644202; PubMed=11707072;
 RA Raymond A., Friedli M., Neergard Henriksen C., Chapot F., Deutsch S., Ucla C., Roessier C., Lyle R., Guipponi M., Antonarakis S.E.;
 RT "From PREDs and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map."; Genomics 78:46-54(2001).
 RL [6]
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- TISSUE SPECIFICITY: Found in spleen, testis, prostate and fetal liver. Expression limited to vascular muscle of testis, smooth muscle of prostate stroma, heart muscle, skeletal muscle, crypts of small intestine, and red pulp of spleen.
 CC -1- PTM: N-glycosylated.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

 CC EMBL; AF257472; AAL05981.1; -;
 DR EMBL; AL163217; CAB90388.1; -;
 DR EMBL; BC009418; AAB09418.1; -;
 DR EMBL; AK022689; BAB14181.1; ALT_INIT.
 DR HSSP; P22897; IEGG.
 DR Genew; HGNC:17807; CHODL.
 DR MIM; 607247; -;
 DR InterPro; IPR001304; Lectin_C.

DR Pfam; PF00059; lectin_c; 1.
 DR SMART; SM00034; CLECT; 1.
 DR PROSITE; PS00615; C TYPE LECTIN 1; FALSE_NEG.
 DR PROSITE; PS00041; C TYPE LECTIN 2; 1.
 KW Lectin; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 273 CHONDROLECTIN.
 FT DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 217 237 POTENTIAL.
 FT DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 35 179 C-TYPE LECTIN.
 FT CARBOHYD 86 86 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 273 AA; 30431 MW; F4890AFAFB572A311 CRC64;

Query Match 50.4%; Score 561.5; DB 1; Length 273;
 Best Local Similarity 60.1%; Pred. No. 5.7e-43;
 Matches 101; Conservative 25; Mismatches 35; Indels 7; Gaps 3;

QY 4 RLSSGQVCRGGTQRCYKVIYFHTSRRLNFEAEKACRRGGQVLSIESDEQKLIK 63
 Db 23 RVSGQKVCFAFKPCYKMAVFHELSRVSFQEARLACESEGGVLLSEAEQKLIK 82
 QY 64 FIENLLP-----SDGFTWIGLARREKQSNSTACODLYAWTDGTSIQFRNYYVDPSGCS 118
 Db 83 MLQNLTKPGTGISDGFWIGLWRNGDQT-SGACPDLYQWSDGNSQYRNWYTDPSGCS 141
 QY 119 EVCVVMYHOPAPAGIGGPFQWDDRCNMKNFKICKYSDE-KPAPV 165
 Db 142 EKCVMYHQPTANPLGGPFLYQWDDRCNMKNHYICKYPEINPTAP 189

RESULT 3
 MANR HUMAN
 ID AC MANR HUMAN STANDARD; PRT; 1456 AA.
 IC P22897;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Macrophage mannose receptor precursor (MMR) (CD206 antigen).
 GN MRC1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Placenta;
 RX MEDLINE=90324192; PubMed=2373685;
 RA Taylor M.E., Conary J.T., Lennartz M.R., Stahl P.D., Drickamer K.;
 RT "Primary structure of the mannose receptor contains multiple motifs resembling carbohydrate-recognition domains.";
 RL J. Biol. Chem. 265:12156-12162(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93052405; PubMed=1294118;
 RA Kim S.J., Ruiz N., Bezouska K., Drickamer K.;
 RT "Organization of the gene encoding the human macrophage mannose receptor (MRC1).";
 RL Genomics 14:721-727(1992).
 RN [3]
 RP STUDIES ON THE BINDING OF INDIVIDUAL LECTIN DOMAINS.
 RX MEDLINE=92112893; PubMed=1730714;
 RA Taylor M.E., Bezouska K., Drickamer K.;
 RT "Contribution to ligand binding by multiple carbohydrate-recognition domains in the macrophage mannose receptor.";
 RL J. Biol. Chem. 267:1719-1726(1992).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 642-788.
 RX MEDLINE=20347275; PubMed=10779515;
 RA Feinberg H., Park-Snyder S., Kolatkar A.R., Heise C.T., Taylor M.E., Weis W.I.;
 RT "Structure of a C-type carbohydrate recognition domain from the macrophage mannose receptor.";

RL J. Biol. Chem. 275:21539-21548(2000).
 CC -!- FUNCTION: MEDIATES THE ENDOCYTOSIS OF GLYCOPROTEINS BY
 CC MACROPHAGES, IN SEVERAL RECOGNITION AND UPTAKE PROCESSES.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- MISCELLANEOUS: CRDS 1-3 HAVE AT MOST VERY WEAK AFFINITY FOR
 CC CARBOHYDRATE. CRD 4 SHOWS THE HIGHEST AFFINITY BINDING AND HAS
 CC MULTISPECIFICITY FOR A VARIETY OF MONOSACCHARIDES. AT LEAST 3 CRDS
 CC (4, 5, AND 7) ARE REQUIRED FOR HIGH AFFINITY BINDING AND
 CC ENDOCYTOSIS OF MULTIVALENT GLYCOCONGUGATES.
 CC -!- SIMILARITY: Contains 8 C-type lectin family domains.
 CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC -!- DATABASE: NAME=PROW; NOTE=PROW 2:85-89(2001);
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1644341535.g.htm".

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).

EMBL; J05550; AAA59868.1; -
 EMBL; M93221; AAA60389.1; -
 EMBL; M93192; AAA60389.1; JOINED.
 EMBL; M93193; AAA60389.1; JOINED.
 EMBL; M93194; AAA60389.1; JOINED.
 EMBL; M93195; AAA60389.1; JOINED.
 EMBL; M93196; AAA60389.1; JOINED.
 EMBL; M93197; AAA60389.1; JOINED.
 EMBL; M93198; AAA60389.1; JOINED.
 EMBL; M93199; AAA60389.1; JOINED.
 EMBL; M93200; AAA60389.1; JOINED.
 EMBL; M93201; AAA60389.1; JOINED.
 EMBL; M93202; AAA60389.1; JOINED.
 EMBL; M93203; AAA60389.1; JOINED.
 EMBL; M93204; AAA60389.1; JOINED.
 EMBL; M93205; AAA60389.1; JOINED.
 EMBL; M93206; AAA60389.1; JOINED.
 EMBL; M93207; AAA60389.1; JOINED.
 EMBL; M93208; AAA60389.1; JOINED.
 EMBL; M93209; AAA60389.1; JOINED.
 EMBL; M93210; AAA60389.1; JOINED.
 EMBL; M93211; AAA60389.1; JOINED.
 EMBL; M93212; AAA60389.1; JOINED.
 EMBL; M93213; AAA60389.1; JOINED.
 EMBL; M93214; AAA60389.1; JOINED.
 EMBL; M93215; AAA60389.1; JOINED.
 EMBL; M93216; AAA60389.1; JOINED.
 EMBL; M93217; AAA60389.1; JOINED.
 EMBL; M93218; AAA60389.1; JOINED.
 EMBL; M93219; AAA60389.1; JOINED.
 EMBL; M93220; AAA60389.1; JOINED.
 PIR; A36563; A36563.
 PDB; 1EGG; 30-AUG-00.
 PDB; 1EGI; 30-AUG-00.
 Genew; HGNC:7228; MRC1.
 MIM; 153618; -
 GO; GO:0005887; C: integral to plasma membrane; TAS.
 GO; GO:0005537; F: mannose binding activity; TAS.
 GO; GO:0004872; F: receptor activity; TAS.
 GO; GO:0006898; P: receptor mediated endocytosis; TAS.
 InterPro; IPR002353; AntifreezeII.
 InterPro; IPR000562; FN_Type_II.
 InterPro; IPR001304; Lectin_C.
 InterPro; IPR000772; Ricin_B_lectin.
 Pfam; PF00040; fn2; 1.
 Pfam; PF00059; lectin_c; 8.
 Pfam; PF00652; Ricin_B_lectin; 2.
 PRINTS; PR00013; FNTYPEII.
 PRINTS; PR00356; ANTIFREEZEII.
 ProDom; PD000995; FN_Type_II; 1.
 SMART; SM00034; CLECT; 8.

DR SMART; SMO0059; FN2; 1.
 DR PROSITE; PS00458; RICIN; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 6.
 DR PROSITE; PS00411; C-TYPE LECTIN 2; 8.
 DR PROSITE; PS00023; FIBRONECTIN 2; 1.
 DR PROSITE; PS00231; RICIN B LECTIN; 1.
 KW Receptor; Signal; Calcium-binding; Transmembrane; Repeat;
 KW Glycoprotein; Endocytosis; Macrophage; Lectin; Antigen; 3D-structure.
 FT SIGNAL 1 18
 FT CHAIN 19 1456
 FT DOMAIN 19 1383
 FT TRANSMEM 1384 1411
 FT DOMAIN 1412 1456
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 22 142
 FT DOMAIN 157 212
 FT DOMAIN 216 344
 FT DOMAIN 360 490
 FT DOMAIN 502 629
 FT DOMAIN 644 781
 FT DOMAIN 805 926
 FT DOMAIN 943 1083
 FT DOMAIN 1100 1216
 FT DOMAIN 1228 1359
 FT DISULFID 646 659
 FT DISULFID 680 777
 FT DISULFID 753 769
 FT CARBOHYD 104 104
 FT CARBOHYD 344 344
 FT CARBOHYD 529 529
 FT CARBOHYD 526 526
 FT CARBOHYD 930 930
 FT CARBOHYD 1160 1160
 FT CARBOHYD 1205 1205
 FT CARBOHYD 1311 1311
 FT TURN 648 649
 FT STRAND 652 652
 FT TURN 654 655
 FT STRAND 658 663
 FT HELIX 667 669
 FT STRAND 671 671
 FT HELIX 673 683
 FT TURN 684 684
 FT STRAND 686 687
 FT HELIX 693 705
 FT TURN 706 707
 FT STRAND 709 710
 FT TURN 712 718
 FT STRAND 723 724
 FT TURN 727 727
 FT TURN 729 730
 FT STRAND 733 733
 FT TURN 741 742
 FT HELIX 746 748
 FT STRAND 752 757
 FT TURN 758 761
 FT STRAND 764 768
 FT TURN 769 770
 FT STRAND 773 780
 FT TURN 781 782
 SQ SEQUENCE 1456 AA; 166011 MW; 2645EAF3C576A5E3 CRC64;

Query Match 16.3%; Score 182; DB 1; Length 1456;
 Best Local Similarity 25.5%; Pred. No. 3.4e-08;
 Matches 52; Conservative 37; Mismatches 61; Indels 54; Gaps 8;

QY 21 KYVYFHDTSRLNFEENKACRDGQGLVSIQSEDEQKLEKFIENLLPSDGDFTWGLR 80
 DB 807 KYDYQYFSEKETMDNARAFCKNFGLVSIQSESEKFLWKYV-NRNDQAQSAFYIGLL 865
 QY 81 RREKQSNACQDLVAWTDGSIQFRNYYVDEPSCGS--EVCVMYVHQPSAPAGIGPY 138
 DB 866 ISLDKK-----FAWMDGSKVDYVSWATGEFNFANEDENCVTMT-----SNSGF---- 908

QY 139 MFQWDDRCNMKNPFICKYDEK---PAVPSRAEGEETELTTPVLPEETOE----- 187
 DB 909 ---WNDINCYPNAPICQHRHSSINATVMP-----TPDSVSGCKEGNIFYSN 954
 QY 188 -----EDAKYTFKESREAAAL 202
 DB 955 KCFKIFGFMEEERKNWQEARAKCI 978
 RESULT 4
 PGCN MOUSE
 ID PGCN MOUSE STANDARD; PRT; 1268 AA.
 AC P55066;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 15-SEP-2003 (Rel. 42, Last annotation update)
 GN Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
 OS CSPG3 OR NCAN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Brain;
 RX MEDLINE=96039250; PubMed=7490074;
 RA Rauch U., Grimpe B., Kulbe G., Arnold-Ammer I., Beier D.,
 RT Faessler R.;
 "Structure and chromosomal localization of the mouse neurocan gene.";
 RL Genomics 28:405-410(1995).
 CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
 development by binding to neural cell adhesion molecules (NG-CAM
 and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
 acid.
 CC -!- TISSUE SPECIFICITY: BRAIN.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see http://www.iab-sib.ch/announce/
 or send an email to license@sib-sib.ch).
 DR EMBL; X84727; CAA59216.1; -.
 DR PIR; S52781; S52781.
 DR HSP; P00740; LEDM.
 DR MGD; MG1:104694; Cspg3.
 DR InterPro; IPR002353; AntifreezeII.
 DR InterPro; IPR000152; Asx hydroxyl.
 DR InterPro; IPR000742; EGF-2.
 DR InterPro; IPR001881; EGF-Ca.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003006; Ig MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00059; Lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.
 DR PRINTS; PR01265; LINKMODULE.
 DR PRINTS; PR00356; ANTIFREEZEII.
 DR ProDom; PD000918; Link; 2.


```

DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG_1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1268
FT DOMAIN 37 157
FT DOMAIN 158 253
FT DOMAIN 259 355
FT DOMAIN 960 996
FT DOMAIN 998 1034
FT DOMAIN 1036 1165
FT DOMAIN 1166 1224
FT DISULFID 58 139
FT DISULFID 181 252
FT DISULFID 205 226
FT DISULFID 279 354
FT DISULFID 303 324
FT DISULFID 964 975
FT DISULFID 969 984
FT DISULFID 986 995
FT DISULFID 1040 1051
FT DISULFID 1068 1160
FT DISULFID 1136 1152
FT DISULFID 1167 1210
FT DISULFID 1196 1223
FT CARBOHYD 121 121
FT CARBOHYD 339 339
FT CARBOHYD 742 742
FT CARBOHYD 978 978
FT CARBOHYD 1175 1175
FT CARBOHYD 1175 1175
SQ SEQUENCE 1268 AA; 137200 MW; 3014E8E202A2FAEC CRC64;

Query Match 15.9%; Score 177; DB 1; Length 1268;
Best Local Similarity 31.5%; Pred. No. 8.1e-08;
Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;

Qy 17 QRPCKYVIYFHTSRRLNFEAKACRRDGGQVLSIESDEQKLIKFIENLLPSDGDFF 76
Db 1048 QGHCYR--YF---AHRRAWEADAERDCRRAGHLTSVHSPEHKFINSF-----GHENSW 1096

Qy 77 IGLRRREKQSNSTACQDLVATWDTGSIQFNNVTVDEPS---CGSEVCVVMYHQPSPAPAG 133
Db 1097 IGLNDRTVRD-----FQWTDNTGLQVENWREKQPDNFFAGGEDCVVMVAHESG--- 1145

Qy 134 IGGPYMFQWQNDRCNNKNPFCK 156
Db 1146 -----RWNDVPCYNLPVCK 1161

RESULT 5
PGCN HUMAN
ID PGCN HUMAN STANDARD; PRT; 1321 AA.
AC O1459; Q9UPK6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
GN CSPG3 OR NCAN OR NEUR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99013874; PubMed=9795216;
RA Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Lennon G.G.;
RT "Characterization of the human neurocan gene, CSPG3.";
RL Gene 221:199-205(1998).
RN [2]
RP SEQUENCE OF 1-990 AND 1007-1321 FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Gaines J.,
RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Cosfield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of an -1 Mb region containing the MEF2B gene in
RL 19p12.";
RP Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC development by binding to neural cell adhesion molecules (NG-CAM
CC and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC acid.
CC -!- TISSUE SPECIFICITY: Brain.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF026547; AAC80576.1; -
DR EMBL; AC003110; AAB86655.1; -
DR EMBL; AC005254; AAC25581.1; -
DR HSSP; P00740; 1EDM.
DR MIM; 600826; -.
DR Genew; HGNC:2465; CSPG3.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Like.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; XLink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00409; IG_1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.

```

```
DR PROSITE; PS01187; EGF CA; 1.
DR PROSITE; PS0835; IG LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1321 NEUCAN CORE PROTEIN.
FT DOMAIN 38 153 IG-LIKE V-TYPE.
FT DOMAIN 159 254 LINK 1.
FT DOMAIN 260 356 LINK 2.
FT DOMAIN 1008 1044 EGF-LIKE 1.
FT DOMAIN 1046 1082 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1084 1213 C-TYPE LECTIN.
FT DOMAIN 1214 1272 SUSHI.
FT DISULFID 59 140 BY SIMILARITY.
FT DISULFID 182 253 BY SIMILARITY.
FT DISULFID 206 227 BY SIMILARITY.
FT DISULFID 280 355 BY SIMILARITY.
FT DISULFID 304 325 BY SIMILARITY.
FT DISULFID 1012 1032 BY SIMILARITY.
FT DISULFID 1017 1032 BY SIMILARITY.
FT DISULFID 1034 1043 BY SIMILARITY.
FT DISULFID 1088 1099 BY SIMILARITY.
FT DISULFID 1116 1208 BY SIMILARITY.
FT DISULFID 1184 1200 BY SIMILARITY.
FT DISULFID 1215 1258 BY SIMILARITY.
FT DISULFID 1244 1271 BY SIMILARITY.
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1026 1026 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1223 1223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1234 1234 N -> Y (IN REF. 2).
FT CONFLICT 1254 1254 V -> A (IN REF. 2).
FT CONFLICT 1282 1282 G -> R (IN REF. 2).
SQ SEQUENCE 1321 AA; 142972 MW; 2EF47F823DB98088 CRC64;

Query Match 15.8%; Score 176; DB 1; Length 1321;
Best Local Similarity 31.5%; Pred. No. 1e-07;
Matches 45; Conservative 18; Mismatches 48; Indels 32; Gaps 6;

QY 17 QRPCKYVYFHTSRRLNFEAKACRRDGGQLVSEDEQKLIKFKFLENLPDGDWF 76
Db 1096 QGHCYR--YF---AHRRAWEDAEDKDCRRSGHLTSVHSPEEHSFINSF-----GHENTW 1144
QY 77 IGLRREEKQNSTACQDLYAWTDGSIQFRNMYVDPEPS---CGSVCVVMYHQPSAPAG 133
Db 1145 IGLNDRIVERD-----FQWTDNTGLQFENWENQPNFAGGEDCVVMVAHESG---- 1193
QY 134 IGGPYMFQWDRRCNNMKNPFCK 156
Db 1194 -----RWNDVPCNYNLPYVCK 1209

RESULT 6
PGCV_BOVIN
ID PGCV_BOVIN STANDARD; PRT; 3381 AA.
AC P81282; O77609; O77610; O77611; O77612;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia)
DE hyaluronate-binding protein) (GHAP).
GN CPBG2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS V0; V1; V2 AND V3).
RC TISSUE=Forebrain;
RX MEDLINE=98288320; PubMed=9624174;
```

| FT | VARSP | PLIC | 350 | 1336 | /FTid=VSP_003078. Missing (in isoform V1). /FTid=VSP_003079. Missing (in isoform V2). /FTid=VSP_003080. Missing (in isoform V3). /FTid=VSP_003081. MISSING (IN REF. 2). MISSING (IN REF. 2). N -> D (IN REF. 2). Q -> D (IN REF. 2). C -> R (IN REF. 2). |
|---|---|--|------------|-----------------|---|
| FT | VARSP | PLIC | 1337 | 3074 | |
| FT | VARSP | PLIC | 350 | 3074 | |
| FT | CONFLICT | 25 | 25 | | |
| FT | CONFLICT | 51 | 51 | | |
| FT | CONFLICT | 89 | 89 | | |
| FT | CONFLICT | 96 | 96 | | |
| FT | CONFLICT | 346 | 346 | | |
| SEQ | SEQUENCE | 3381 AA; | 369984 MW; | F09716FA778D459 | CRC64; |
| Query Match | | | | | |
| Best Local Similarity 15.7%; Score 174.5; DB 1; Length 3381; | | | | | |
| Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11; | | | | | |
| Qy | 17 | QPCYKVIYPHDSRLNFEFEAKACRRDGGQVLSIESDEQKLIKFIENLLPSDGF- | 75 | | |
| Db | 3162 | QGQCYK-YF--AHRTWDAERECRLQGAHLTSLSHEEQMFVNRV-----CHDQV | 3209 | | |
| Qy | 76 | WIGLRREEKQSNSTACQDLYAWTDGSIQFRWYVDEP-----SCGEVGVVWYHPSAP | 131 | | |
| Db | 3210 | WIGL-----NDKMFEDHDFRDTGSLTQYENWRPNQDPSFSTGDCVVIWHENG-- | 3259 | | |
| Qy | 132 | AGIGGPMFOWDDRCNMKNFKICKYS---DEKPAVPREAEGE----- | 172 | | |
| Db | 3260 | -----QWMDVPCNYHLTYTCRKGTVACGPPVVENAKTFGKMKPYEINSLIRYHC | 3310 | | |
| Qy | 173 | -----ETELTT-----PVL-----PEETQBEDAKKTFKESREAAALN | 203 | | |
| Db | 3311 | KDGFQIRHLPTIRCLNGRWAMPKITCLNPSAYQRTYSKKYFKNSSAKDN | 3361 | | |
| RESULT 7 | | | | | |
| PGCN | RAT | STANDARD; | PRT; | 1257 | AA. |
| AC | PC5067; | | | | |
| DT | 01-OCT-1996 | (Rel. 34, Created) | | | |
| DT | 01-OCT-1996 | (Rel. 34, Last sequence update) | | | |
| DT | 15-SEP-2003 | (Rel. 42, Last annotation update) | | | |
| DE | Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3) | | | | |
| DE | (245 kDa early postnatal core glycoprotein) (Contains: 150 kDa adult | | | | |
| GN | core glycoprotein). | | | | |
| GN | CSPG3 OR NCAN. | | | | |
| OS | Rattus norvegicus (Rat). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | | | | |
| OX | NCBI_TaxID=10116; | | | | |
| RP | SEQUENCE FROM N.A., AND PARTIAL SEQUENCE. | | | | |
| RP | STRAIN=Sprague-Dawley; TISSUE=Brain; | | | | |
| RC | MEDLINE=92406907; PubMed=1326557; | | | | |
| RA | Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.; | | | | |
| RT | "Cloning and primary structure of neurocan, a developmentally | | | | |
| RT | regulated, aggregating chondroitin sulfate proteoglycan of brain." | | | | |
| RL | J. Biol. Chem. 267:19536-19547(1992). | | | | |
| RN | [2] | | | | |
| RP | CHARACTERIZATION. | | | | |
| RP | MEDLINE=94230574; PubMed=7513709; | | | | |
| RX | Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.; | | | | |
| RA | Margolis R.U., Grumet M.; | | | | |
| RT | "The neuronal chondroitin sulfate proteoglycan neurocan binds to the | | | | |
| RT | neuronal cell adhesion molecules Ng-CAM/Li/NILE and N-CAM, and inhibits | | | | |
| RT | neuronal adhesion and neurite outgrowth." | | | | |
| RL | J. Cell Biol. 125:669-680(1994). | | | | |
| CC | -!- FUNCTION: May modulate neuronal adhesion and neurite growth during | | | | |
| CC | development by binding to neural cell adhesion molecules (NG-CAM | | | | |
| CC | and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic | | | | |
| CC | acid. | | | | |
| CC | -!- TISSUE SPECIFICITY: EARLY POSTNATAL AND ADULT BRAIN; NOT EXPRESSED | | | | |
| CC | IN KIDNEY, LUNG, LIVER AND MUSCLE. | | | | |

CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -1- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
CC DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL
CC BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M97161; AAC37679.1; --
CC PIR; S28764; S28764.
CC HSSP; P00740; 1EDM.
CC
CC InterPro; IPR002353; Antifreeze1.
CC InterPro; IPR000152; Asx_Hydroxyl.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_CA.
CC InterPro; IPR006209; EGF_Like.
CC InterPro; IPR007110; Ig_Like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF_2.
CC Pfam; PF00047; Ig_1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; xlink; 2.
CC PRINTS; PR01265; LINKMODULE.
CC PRINTS; PR00356; ANTIFREEZE1.
CC ProDom; PD000918; Link; 2.
CC SMART; SM00032; CCF; 1.
CC SMART; SM00034; CLECT; 1.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00409; IG; 1.
CC SMART; SM00445; LINK; 2.
CC PROSITE; PS00010; ASX_HYDROXYL; 1.
CC PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
CC PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
CC PROSITE; PS00022; EGF_1; 3.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS01187; EGF_CA; 1.
CC PROSITE; PS00835; IG_Like; 1.
CC PROSITE; PS01241; Link; 2.
CC Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
CC EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
CC SIGNAL 1 22
CC CHAIN 23 1257 NEUROCAN CORE PROTEIN.
CC CHAIN 639 1257 150 kDa ADULT CORE GLYCOPROTEIN.
CC DOMAIN 37 157 IG-LIKE V-TYPE.
CC DOMAIN 158 253 LINK 1.
CC DOMAIN 259 355 LINK 2.
CC DOMAIN 949 985 EGF-LIKE 1.
CC DOMAIN 987 1023 EGF-LIKE 2.
CC DOMAIN 1025 1154 C-TYPE LECTIN.
CC DOMAIN 1155 1213 SUSHI.
CC DISULFID 58 139 BY SIMILARITY.
CC DISULFID 181 252 BY SIMILARITY.
CC DISULFID 205 226 BY SIMILARITY.
CC DISULFID 279 354 BY SIMILARITY.
CC DISULFID 303 324 BY SIMILARITY.
CC DISULFID 953 964 BY SIMILARITY.

FT DISULFID 958 973 BY SIMILARITY.
FT DISULFID 975 984 BY SIMILARITY.
FT DISULFID 1029 1040 BY SIMILARITY.
FT DISULFID 1057 1149 BY SIMILARITY.
FT DISULFID 1125 1141 BY SIMILARITY.
FT DISULFID 1156 1199 BY SIMILARITY.
FT DISULFID 1185 1212 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 944 944 O-LINKED (XYL. .) (POTENTIAL).
FT CARBOHYD 967 967 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1164 1164 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1257 AA; 135544 MW; 992B33DCFA19EE1B CRC64;
Query Match 15.6%; Score 174; DB 1; Length 1257;
Best Local Similarity 30.8%; Pred. No. 1.5e-07;
Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;
QY 17 QRPCYKVIYFHDTSRLNFEAKACRRDGGQLVSESEDEQKLIKFIENLLPSDGDWF 76
DB 1037 QGHYR--YF--AHRAWEDAERDCRRAGHLTSVHSPEHKFINSF-----GHNSW 1085
QY 77 IGLRREEKQSNSTACQDLYAWTGDSTISQFRNMYVDEPS---CGSEVYVVMYHQPSAPAG 133
DB 1086 IGLNDRTVRD-----FQMTDNTGLQYENWREKQPDNFFAGGDCVVMVAHENG--- 1134
QY 134 IGGPYMFQWDDRCNMKNPFICK 156
DB 1135 -----RMNDVPCNYPYVCK 1150
RESULT 8
PGCV_RAT STANDARD; PRT; 2738 AA.
AC Q9ERB4; O08592; O08564; Q9RIK4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP) (Fragments).
GN CP5G2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.
RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wight T.N.;
RA "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [2]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=98308094; PubMed=9642104;
RA Milev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,
RA Margolis R.K., Margolis R.U.;
RT "Differential regulation of expression of hyaluronan-binding
RT proteoglycans in developing brain: aggrecan, versican, neurocan, and
RT brevican.";
RL Biochem. Biophys. Res. Commun. 247:207-212(1998).
RN [3]
RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).
RC TISSUE=Kidney;
RX MEDLINE=98094159; PubMed=9434070;
RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RA "Proteoglycan expression in the normal rat kidney.";
RL Nephron 77:461-470(1997).

| | |
|----|---|
| KW | Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain; |
| KW | Hyaluronic acid; Alternative splicing. |
| FT | POTENTIAL. |
| FT | VERSICAN CORE PROTEIN. |
| FT | IG-LIKE V-TYPE. |
| FT | LINK 1. |
| FT | LINK 2. |
| FT | GAG-ALPHA |
| FT | (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN). |
| FT | GAG-BETA. |
| FT | EGF-LIKE 1. |
| FT | EGF-LIKE 2, CALCIUM-BINDING. |
| FT | C-TYPE LECTIN. |
| FT | SUSHI. |
| FT | BY SIMILARITY. |
| FT | BY SIMILARITY. |
| FT | BY SIMILARITY. |
| FT | BY SIMILARITY. |
| FT | BY SIMILARITY. |
| FT | BY SIMILARITY. |
| FT | BY SIMILARITY. |
| FT | BY SIMILARITY. |
| FT | BY SIMILARITY. |
| FT | BY SIMILARITY. |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | Misling (in isoform V3). |
| FT | /FTID=VSP 003091. |
| FT | PRAYOTYSKYLVKNSSVKNSINTSKHEHRSRWQETR |
| FT | R -> RKWSFRKNGPCFNKY (in isoform Vint). |
| FT | /FTID=VSP 003092. |
| FT | AEREC -> NSARG (in REF. 4). |
| FT | SEQUENCE 2738 AA; 300004 MW; 12CA626D58BD8C6A CRC64; |
| SQ | Query Match Best Local Similarity 28.5%; Pred No. 3.9e-07; Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8; |
| QY | 17 QRPCKVIYFHDTSRRLNFEAEAKCRDRGGOLVSIESEDEOKLIEKFENLLPSDGDF- 75 |
| Db | 2519 QCQCXYK-YF--AHRRRTWDAAERECRLQGAHLTSLSHEEQMFVNRV-----GHDIQ 2566 |
| QY | 76 WGLRREKQSNSTACQDLIAWTGDSISOFNWTVDEP-----SCGEVCVVVMHQPSAP 131 |
| Db | 2567 WGL-----NDKMFEHDFFWTDGSALQYENWRPNQDPSPFSAGEPCVIIWHENG-- 2616 |
| QY | 132 AGTGGPYFMQWDDRCNNKNNFIKCYK-----DEKPAVPISRBAE 172 |
| Db | 2617 -----QWNVPCNVHLTTCCKGTGACQPPIVENAKTFGK 2652 |

```

RESULT 9
PCV_MOUSE
ID PCV_MOUSE STANDARD; PRT: 3358 AA.
AC Q62059; Q62058; Q6CU00;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE Chondroitin sulfate proteoglycan core protein 2 (PG-M).
DE CSF2.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM V0; V1 AND V2).
RC STRAIN=C57BL/6; and Swiss Webster; TISSUE=Brain;
RX MEDLINE=95122551; PubMed=7822336;
RA Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;
RT "Multiple forms of mouse PG-M, a large chondroitin sulfate
RT proteoglycan generated by alternative splicing.";
RL J. Biol. Chem. 270:958-965(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM V3).
RC STRAIN=C57BL/6;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
RL J. Biol. Chem. 270:3914-3918(1995).
RN [3]
RP SEQUENCE OF 1-1692 FROM N.A. (ISOFORM V1).
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=21085660; PubMed=1217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Ronald M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [4]
RP INTERACTION WITH FBLN1.
RX PubMed=10400671;
RA Aspegberg A., Adam S., Kostka G., Timpl R., Heinegaard D.;
RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and
RT versican.";
RL J. Biol. Chem. 274:20444-20449(1999).
CC -1- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -1- SUBUNIT: Interacts with FBLN1.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=Q62059-1; Sequence=Displayed;

```

```

CC CC Name=V1;
CC CC IsoId=Q62059-2; Sequence=VSP_003087, VSP_003088;
CC CC Name=V2;
CC CC IsoId=Q62059-3; Sequence=VSP_003089;
CC CC Name=V3;
CC CC IsoId=Q62059-4; Sequence=VSP_003087, VSP_003090;
CC CC TISSUE SPECIFICITY: V2 is found only in brain.
CC CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC CC -1- SIMILARITY: Contains 2 link domains.
CC CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; D16263; BAA03796.1; -
CC CC EMBL; D28599; -; NOT ANNOTATED_CDS.
CC CC EMBL; D32040; BAA06802.1; -
CC CC EMBL; AK014525; BAB29411.1; -
CC CC HSSP; P01132; IEPG.
CC CC MGD; MGI:102889; Cspg2.
CC CC InterPro; IPR000152; Asx_hydroxyl.
CC CC InterPro; IPR000742; EGF_2.
CC CC InterPro; IPR001881; EGF_Ca.
CC CC InterPro; IPR006209; EGF_Like.
CC CC InterPro; IPR007110; Ig-Like.
CC CC InterPro; IPR003599; Ig.
CC CC InterPro; IPR003006; Ig_MHC.
CC CC InterPro; IPR001304; Lectin_C.
CC CC InterPro; IPR000538; Link.
CC CC InterPro; IPR000436; Sushi_SCR_CCP.
CC CC Pfam; PF00008; EGF; 2.
CC CC Pfam; PF00047; Ig; 1.
CC CC Pfam; PF00059; lectin_c; 1.
CC CC Pfam; PF00084; sushi_1.
CC CC Pfam; PF00193; Xlink; 2.
CC CC PRINTS; PR01265; LINKMODULE.
CC CC ProDom; PD000918; Link; 2.
CC CC SMART; SM00032; CCP; 1.
CC CC SMART; SM00034; CLECT; 1.
CC CC SMART; SM00179; EGF_CA; 1.
CC CC SMART; SM00409; IG_1.
CC CC SMART; SM00445; LINK; 2.
CC CC PROSITE; PS00010; ASX_HYDROXYL; 1.
CC CC PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
CC CC PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
CC CC PROSITE; PS00022; EGF_1_2.
CC CC PROSITE; PS01186; EGF_2; 1.
CC CC PROSITE; PS01187; EGF_CA; 1.
CC CC PROSITE; PS00835; IG_Like; 1.
CC CC PROSITE; PS01241; LINK; 2.
CC CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
CC Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
CC Hyaluronic acid; Alternative splicing.
CC SIGNAL 1 20 POTENTIAL.
CC FT CHAIN 21 3358 VERSICAN CORE PROTEIN.
CC FT DOMAIN 21 146 IG-LIKE V-TYPE.
CC FT DOMAIN 167 244 LINK 1.
CC FT DOMAIN 265 346 LINK 2.
CC FT DOMAIN 348 1308 GAG-ALPHA
CC FT (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
CC FT DOMAIN 1309 3052 GAG-BETA.
CC FT DOMAIN 3052 3088 EGF-LIKE 1.
CC FT DOMAIN 3090 3126 EGF-LIKE 2.
CC FT DOMAIN 3139 3253 C-TYPE LECTIN.

```

```

DE  DE hyaluronate-binding protein) (GHAP).
GN  CSPG2.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN  [1]_TaxID=9606;
RP  SEQUENCE FROM N.A. (ISOFORM V0).
RX  MEDLINE=95105188; PubMed=7528742;
RA  Naso M.F., Zimmermann D.R., Iozzo R.V.;
RT  "Characterization of the complete genomic structure of the human
RL  versican gene and functional analysis of its promoter.";
RN  [2]
RP  SEQUENCE FROM N.A. (ISOFORM V1).
RX  TISSUE=Placenta;
RA  Zimmermann D.R., Ruoslahti E.;
RT  "Multiple domains of the large fibroblast proteoglycan, versican.";
RN  [3]
RP  SEQUENCE FROM N.A. (ISOFORM V2).
RX  TISSUE=Glial tumor;
RA  Dours-Zimmermann M.T., Zimmermann D.R.;
RT  "A novel glycosaminoglycan attachment domain identified in two
RL  alternative splice variants of human versican.";
RN  [4]
RP  SEQUENCE OF 2711-3396 FROM N.A.
RX  TISSUE=Lung fibroblast;
RA  MEDLINE=86007514; PubMed=2820964;
RT  "A fibroblast chondroitin sulfate proteoglycan core protein contains
RL  lectin-like and growth factor-like sequences.";
RN  [5]
RP  SEQUENCE OF 251-347 FROM N.A.
RX  MEDLINE=9312792; PubMed=1478664;
RA  Iozzo R.V., Naso M.F., Cannizzaro L.A., Waasmuth J.J.;
RT  "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
RL  human chromosome 5 (5q12-5q14)";
RN  [6]
RP  SEQUENCE FROM N.A. (ISOFORM V3).
RX  TISSUE=Brain;
RA  MEDLINE=95181355; PubMed=7876137;
RT  "Expression of PG-M(V3), an alternatively spliced form of PG-M
RL  without a chondroitin sulfate attachment in region in mouse and human
RN  tissues.";
RN  [7]
RP  SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RX  TISSUE=Aortic smooth muscle;
RA  MEDLINE=93327053; PubMed=10397680;
RT  "Versican/PG-M isoforms in vascular smooth muscle cells.";
RN  [8]
RP  PARTIAL SEQUENCE.
RX  TISSUE=Brain;
RA  MEDLINE=89174663; PubMed=2466833;
RT  "Isolation and partial characterization of a glial
RL  hyaluronate-binding protein.";
RN  [9]
RP  TISSUE SPECIFICITY OF ISOFORMS.
RX  MEDLINE=96213482; PubMed=8627343;

```

```

DE  DE hyaluronate-binding protein) (GHAP).
GN  CSPG2.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN  [1]_TaxID=9606;
RP  SEQUENCE FROM N.A. (ISOFORM V0).
RX  MEDLINE=95105188; PubMed=7528742;
RA  Naso M.F., Zimmermann D.R., Iozzo R.V.;
RT  "Characterization of the complete genomic structure of the human
RL  versican gene and functional analysis of its promoter.";
RN  [2]
RP  SEQUENCE FROM N.A. (ISOFORM V1).
RX  TISSUE=Placenta;
RA  Zimmermann D.R., Ruoslahti E.;
RT  "Multiple domains of the large fibroblast proteoglycan, versican.";
RN  [3]
RP  SEQUENCE FROM N.A. (ISOFORM V2).
RX  TISSUE=Glial tumor;
RA  Dours-Zimmermann M.T., Zimmermann D.R.;
RT  "A novel glycosaminoglycan attachment domain identified in two
RL  alternative splice variants of human versican.";
RN  [4]
RP  SEQUENCE OF 2711-3396 FROM N.A.
RX  TISSUE=Lung fibroblast;
RA  MEDLINE=86007514; PubMed=2820964;
RT  "A fibroblast chondroitin sulfate proteoglycan core protein contains
RL  lectin-like and growth factor-like sequences.";
RN  [5]
RP  SEQUENCE OF 251-347 FROM N.A.
RX  MEDLINE=9312792; PubMed=1478664;
RA  Iozzo R.V., Naso M.F., Cannizzaro L.A., Waasmuth J.J.;
RT  "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
RL  human chromosome 5 (5q12-5q14)";
RN  [6]
RP  SEQUENCE FROM N.A. (ISOFORM V3).
RX  TISSUE=Brain;
RA  MEDLINE=95181355; PubMed=7876137;
RT  "Expression of PG-M(V3), an alternatively spliced form of PG-M
RL  without a chondroitin sulfate attachment in region in mouse and human
RN  tissues.";
RN  [7]
RP  SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RX  TISSUE=Aortic smooth muscle;
RA  MEDLINE=93327053; PubMed=10397680;
RT  "Versican/PG-M isoforms in vascular smooth muscle cells.";
RN  [8]
RP  PARTIAL SEQUENCE.
RX  TISSUE=Brain;
RA  MEDLINE=89174663; PubMed=2466833;
RT  "Isolation and partial characterization of a glial
RL  hyaluronate-binding protein.";
RN  [9]
RP  TISSUE SPECIFICITY OF ISOFORMS.
RX  MEDLINE=96213482; PubMed=8627343;

```

```

DE  DE hyaluronate-binding protein) (GHAP).
GN  CSPG2.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN  [1]_TaxID=9606;
RP  SEQUENCE FROM N.A. (ISOFORM V0).
RX  MEDLINE=95105188; PubMed=7528742;
RA  Naso M.F., Zimmermann D.R., Iozzo R.V.;
RT  "Characterization of the complete genomic structure of the human
RL  versican gene and functional analysis of its promoter.";
RN  [2]
RP  SEQUENCE FROM N.A. (ISOFORM V1).
RX  TISSUE=Placenta;
RA  Zimmermann D.R., Ruoslahti E.;
RT  "Multiple domains of the large fibroblast proteoglycan, versican.";
RN  [3]
RP  SEQUENCE FROM N.A. (ISOFORM V2).
RX  TISSUE=Glial tumor;
RA  Dours-Zimmermann M.T., Zimmermann D.R.;
RT  "A novel glycosaminoglycan attachment domain identified in two
RL  alternative splice variants of human versican.";
RN  [4]
RP  SEQUENCE OF 2711-3396 FROM N.A.
RX  TISSUE=Lung fibroblast;
RA  MEDLINE=86007514; PubMed=2820964;
RT  "A fibroblast chondroitin sulfate proteoglycan core protein contains
RL  lectin-like and growth factor-like sequences.";
RN  [5]
RP  SEQUENCE OF 251-347 FROM N.A.
RX  MEDLINE=9312792; PubMed=1478664;
RA  Iozzo R.V., Naso M.F., Cannizzaro L.A., Waasmuth J.J.;
RT  "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
RL  human chromosome 5 (5q12-5q14)";
RN  [6]
RP  SEQUENCE FROM N.A. (ISOFORM V3).
RX  TISSUE=Brain;
RA  MEDLINE=95181355; PubMed=7876137;
RT  "Expression of PG-M(V3), an alternatively spliced form of PG-M
RL  without a chondroitin sulfate attachment in region in mouse and human
RN  tissues.";
RN  [7]
RP  SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RX  TISSUE=Aortic smooth muscle;
RA  MEDLINE=93327053; PubMed=10397680;
RT  "Versican/PG-M isoforms in vascular smooth muscle cells.";
RN  [8]
RP  PARTIAL SEQUENCE.
RX  TISSUE=Brain;
RA  MEDLINE=89174663; PubMed=2466833;
RT  "Isolation and partial characterization of a glial
RL  hyaluronate-binding protein.";
RN  [9]
RP  TISSUE SPECIFICITY OF ISOFORMS.
RX  MEDLINE=96213482; PubMed=8627343;

```

RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "Differential expression of versican isoforms in brain tumors.";
 RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
 CC -!- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -!- SUBUNIT: Interacts with PBLN1 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Comment=Additional isoforms seem to exist;
 CC Name=V0;
 CC IsoId=Pl3611-1; Sequence=Displayed;
 CC Name=V1;
 CC IsoId=Pl3611-2; Sequence=VSP_003082, VSP_003083;
 CC Name=V2;
 CC IsoId=Pl3611-3; Sequence=VSP_003084;
 CC Name=V3;
 CC IsoId=Pl3611-4; Sequence=VSP_003082, VSP_003085;
 CC Name=Vint;
 CC IsoId=Pl3611-5; Sequence=VSP_003086;
 CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
 CC in normal brain, gliomas, medulloblastomas, schwannomas,
 CC neurofibromas, and meningiomas; V2 is restricted to normal brain
 CC and gliomas; V3 is found in all these tissues except
 CC medulloblastomas.
 CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; U16306; AAA65018.1; -;
 DR EMBL; X15998; CAA34128.1; -;
 DR EMBL; S52488; AAB24878.1; -;
 DR EMBL; U25555; AAA67565.1; -;
 DR EMBL; D32039; BAA06801.1; -;
 DR EMBL; J02814; AAA36437.1; -;
 DR EMBL; AF084545; AAD48545.1; -;
 DR PIR; S06014; A60979.
 DR HSP; P01132; 1EGF.
 DR Genew; HGNC:2464; CSPG2.
 DR MIM; 118661; -;
 DR GO; GO:0005578; C:extracellular matrix; TAS.
 DR GO; GO:0005204; F:chondroitin sulfate proteoglycan; TAS.
 DR GO; GO:0005540; F:hyaluronic acid binding activity; TAS.
 DR GO; GO:0008037; P:cell recognition; TAS.
 DR GO; GO:0007275; P:development; TAS.
 DR InterPro; IPR000152; Aax_hydroxyl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR007110; Ig-Like.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00059; lectin; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.

DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 DR PROSITE; PSS0041; C-TYPE LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PSS0835; IG_LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW Signal; Repeat; EGF-like domain; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 20
 FT CHAIN 21 3396
 FT DOMAIN 21 146
 FT DOMAIN 167 244
 FT DOMAIN 265 346
 FT DOMAIN 348 1335
 FT DOMAIN 1336 3089
 FT DOMAIN 1089 3125
 FT DOMAIN 3127 3163
 FT DOMAIN 3176 3290
 FT DOMAIN 3295 3353
 FT DISULFID 44 130
 FT DISULFID 172 243
 FT DISULFID 196 217
 FT DISULFID 270 345
 FT DISULFID 294 315
 FT DISULFID 3093 3104
 FT DISULFID 3098 3113
 FT DISULFID 3115 3124
 FT DISULFID 3131 3142
 FT DISULFID 3136 3151
 FT DISULFID 3153 3162
 FT DISULFID 3169 3180
 FT DISULFID 3197 3289
 FT DISULFID 3265 3281
 FT DISULFID 3296 3339
 FT DISULFID 3325 3352
 FT CARBOHYD 57 57
 FT CARBOHYD 330 330
 FT CARBOHYD 615 615
 Query Match 15.6%; Score 174; DB 1; Length 3396;
 Best Local Similarity 28.5%; Pred. No. 5.1e-07;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
 QY 17 QRECYKVIYFHDTSRLNFEAEKACRRDGGQLVSEIEDEQKLIKFIENLLPSDGF- 75
 DB 3177 QGQCYK--YF--AHRTWDAARECLQAHLSILSHEEQMFPNVR-----GHDYQ 3224
 QY 76 WIGLRREKQSNSTACQDLYAWTDGSIQFRNRYVDEP-----SCGSEVCVMYHQPAP 131
 DB 3225 WIGL-----NDKMFHEHDFRWTGDTLQYENRPNQDPFSAGDCVVIWHENG-- 3274
 QY 132 AGIGGYPWFQWDDRCNMKNFKICKYS-----DEKPAVPSREAGE 172
 DB 3275 -----QWNVPCNYHLTYTCCKGTACQGPVVENAKTFGK 3310
 RESULT 11
 ID_PGV_CHICK STANDARD; PRT; 3562 AA.
 AC Q90953; Q90945;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
GN CPSG2
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS V0 AND V1).
RC STRAIN=white leghorn; TISSUE=Limb bud;
RX MEDLINE=93300846; PubMed=8314802;
RA Shinomura T., Nishida Y., Ito K., Kimata K.;
RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
RT expressed during chondrogenesis in chick limb buds. Alternative
RT spliced multi-forms of PG-M and their relationships to versican.";
RL J. Biol. Chem. 268:14461-14469(1993).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=Q09053-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=Q09053-2; Sequence=VSP_003093;
CC -!- TISSUE SPECIFICITY: Prechondrogenic condensation area of
CC developing limb buds.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (By similarity).
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X60226; CAA42787.1; -;
CC EMBL; D13542; BAA02742.1; -;
CC PIR; A47171; A47171.
CC HSSP; P00740; 1EDM.
CC InterPro; IPR000152; Asx hydroxyl.
CC InterPro; IPR000742; EGF2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003006; Ig MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00047; Ig; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 2.
CC PRINTS; PR01265; LINKMODULE.
CC PRODOM; PD000918; Link; 2.
CC SMART; SM00032; CCP; 1.
CC SMART; SM00034; CLECT; 1.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00409; IG; 1.
CC SMART; SM00445; LINK; 2.

DR PROSITE; PS00010; ASX HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS00441; C-TYPE LECTIN_2; 1.
DR PROSITE; PS00022; EGF 1; 2.
DR PROSITE; PS01186; EGF 2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 26
FT CHAIN 27 3562 VERSICAN CORE PROTEIN.
FT DOMAIN 27 143 IG-LIKE V-TYPE.
FT DOMAIN 166 243 LINK 1.
FT DOMAIN 284 345 LINK 2.
FT DOMAIN 3254 3290 EGF-LIKE 1.
FT DOMAIN 3292 3328 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3341 3455 C-TYPE LECTIN.
FT DOMAIN 3460 3518 SUSHI.
FT DISULFID 44 129 BY SIMILARITY.
FT DISULFID 171 242 BY SIMILARITY.
FT DISULFID 195 216 BY SIMILARITY.
FT DISULFID 269 344 BY SIMILARITY.
FT DISULFID 293 314 BY SIMILARITY.
FT DISULFID 3258 3269 BY SIMILARITY.
FT DISULFID 3263 3278 BY SIMILARITY.
FT DISULFID 3280 3289 BY SIMILARITY.
FT DISULFID 3296 3307 BY SIMILARITY.
FT DISULFID 3301 3316 BY SIMILARITY.
FT DISULFID 3318 3327 BY SIMILARITY.
FT DISULFID 3334 3345 BY SIMILARITY.
FT DISULFID 3362 3454 BY SIMILARITY.
FT DISULFID 3430 3446 BY SIMILARITY.
FT DISULFID 3461 3504 BY SIMILARITY.
FT DISULFID 3490 3517 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 235 235 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 709 709 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 948 948 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 1409 1409 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 1479 1479 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 1523 1523 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 1530 1530 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 1625 1625 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 1751 1751 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 1988 1988 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 2088 2088 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 2089 2089 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 2507 2507 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 2642 2642 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 2679 2679 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 2748 2748 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 2762 2762 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 3069 3069 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 3194 3194 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 3232 3232 N-LINKED (GLCNAC... (POTENTIAL).
FT CARBOHYD 3545 3545 N-LINKED (GLCNAC... (POTENTIAL).
FT VARSPLIC 485 1411 Missing (in isoform V1).
FT FTId=VSP_003093.
SQ SEQUENCE 3562 AA; 388078 MW; 9BC5668B8C1602D2 CRC64;

Query Match 15.3%; Score 171; DB 1; Length 3562;
Best Local Similarity 28.5%; Pred. No. 9.9e-07;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

QY 17 QPCKVYIYFHTDTSRLNFEAKACRRDGGQLVSIEDQKLEKFIENLLPSDGF- 75
Db 3342 QGQCYK--YF---AHRRTWDTAERECLQGAHLTSLSHEEQVFNRI-----GHQYQ 3389
QY 76 WIGLRRREKQSNSTACQDLVATDGSISQFRNWTVDPE----SCGSEVCVMVHQPSAP 131

```

Db 3390 WIGL-----NDKMFEDFRWTDGSPLOYNRPNDPSFFSAGEDCVIWHENG-- 3439
Qy 132 AGIGPYMFQWDRDCNWNFKYKYS-----DEKPAVPSRAEGE 172
Db 3440 -----QWNVPCNYHLTYTCKGTGTVACGQPPVVENAKTFGK 3475

RESULT 12
CD93_RAT
ID CD93_RAT STANDARD; PRT; 643 AA.
AC Q9ETG1; Q9J126;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Complement component C1q receptor precursor (Complement component 1, q
DE subcomponent, receptor 1) (C1qR) (C1qR(p)) (C1q/MBL/SPA receptor)
DE (CD93 antigen) (Cell surface antigen AA4).
GN C1QR1 OR CD93 OR C1QRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=PVG; TISSUE=Natural killer cells;
RX MEDLINE=20545218; PubMed=11093152;
RA Lovik G., Vaage J.T., Dissen E., Szpirer C., Ryan J.C., Rolstad B.;
RT "Characterization and molecular cloning of rat C1qR, a receptor on NK
RT cells.";
RL Eur. J. Immunol. 30:3355-3362(2000).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Lung;
RX MEDLINE=20507883; PubMed=10934210;
RA Dean Y.D., McGreal E.P., Akatsu H., Gasque P.;
RT "Molecular and cellular properties of the rat AA4 antigen, a C-type
RT lectin-like receptor with structural homology to thrombomodulin.";
RL J. Biol. Chem. 275:34382-34392(2000).
CC -!- FUNCTION: Receptor (or element of a larger receptor complex) for
CC C1q, mannose-binding lectin (MBL2) and pulmonary surfactant
CC protein A (SPA). May mediate the enhancement of phagocytosis in
CC monocytes and macrophages upon interaction with soluble defense
CC collagens. May play a role in intercellular adhesion.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Widely expressed. Highly expressed in lung and
CC heart. Expressed at lower level in brain, thymus, liver, spleen,
CC intestine, kidney, adrenal gland, muscle and testis. Expressed on
CC circulating natural killer cells.
CC -!- PTM: N- and O-glycosylated (by similarity).
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 5 EGF-like domains.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL; AF136537; AAG01572.1; -.
CC EMBL; AF160978; AAF80402.1; -.
CC HSSP; P35555; 1PMN
CC GO; GO:0016021; C:integral to membrane; ISS.
CC GO; GO:0004872; F:receptor activity; ISS.
CC GO; GO:0016337; P:cell-cell adhesion; ISS.
CC GO; GO:0042116; P:macrophage activation; ISS.
CC GO; GO:0006909; P:phagocytosis; ISS.
CC InterPro; IPR000152; Asx hydroxyl.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF_Like.

```

```

DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00008; EGF; 5.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00010; ASX HYDROXYL; 3.
DR PROSITE; PS00615; C-TYPE LECTIN 1; FALSE_NEG.
DR PROSITE; PSS0041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 3.
DR Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
KW EGF-like domain; Lectin; Glycoprotein.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 643 COMPLEMENT COMPONENT C1Q RECEPTOR.
FT DOMAIN 24 571 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 572 592 POTENTIAL.
FT DOMAIN 593 643 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 31 173 C-TYPE LECTIN.
FT DOMAIN 257 298 EGF-LIKE 1.
FT DOMAIN 299 341 EGF-LIKE 2.
FT DOMAIN 342 381 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 382 423 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 424 462 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 268 282 BY SIMILARITY.
FT DISULFID 284 297 BY SIMILARITY.
FT DISULFID 303 314 BY SIMILARITY.
FT DISULFID 308 325 BY SIMILARITY.
FT DISULFID 327 340 BY SIMILARITY.
FT DISULFID 346 355 BY SIMILARITY.
FT DISULFID 351 364 BY SIMILARITY.
FT DISULFID 366 380 BY SIMILARITY.
FT DISULFID 386 397 BY SIMILARITY.
FT DISULFID 393 406 BY SIMILARITY.
FT DISULFID 408 422 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 433 446 BY SIMILARITY.
FT DISULFID 448 461 BY SIMILARITY.
FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 417 417 E -> K (IN REF. 2).
SQ SEQUENCE 643 AA; 68781 MW; 9AE4C933AD943DB6 CRC64;

Query Match 14.8%; Score 165; DB 1; Length 643;
Best Local Similarity 25.2%; Pred. No. 4.2e-07;
Matches 52; Conservative 35; Mismatches 73; Indels 46; Gaps 10;

Qy 3 GRLLSGQP-----VCRGGTQRPCYKVIYFHDTSRRLNFEAEKAEACRRDGGQLVISED 56
Db 15 GQLWAGAAADSEAVVCEG---TACYTAHW-----GKLSAAEAQHRCNENGNLATVKSSE 66
Qy 57 EQKLEKFIENLL-----PSD---GDFWIGLRRREKQSNSTACODLYAWT-DGSISOFRN 108
Db 67 EARRHVQEAQLLKTAPSETKIGKEWIGLQREKGTCTVHDLPWKGFWSVGGGEDTYSN 126
Qy 109 WY-VDEPSCGSEVCVVMY-----HQPSPAGIGGYPFWQNDRC-----NMKNVF 153
Db 127 WYKASSSCISKRCVKLILDLSLKPHPSHLF-----KWHESPCTPDAPGNSIEGF 177
Qy 154 IKYSDKPAVPSRAEGEBETELTTP 179
Db 178 LCKFNFKGMCSPALOGPGQLTYTTP 203

RESULT 13
PGCA_BOVIN
ID PGCA_BOVIN STANDARD; PRT; 2364 AA.
AC P13608; P79117; Q28159;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSPCP).

```

GN AGC1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_taxid=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Hering T.M., Kollar J., Huynh T.D.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 563-1056 FROM N.A.
RX MEDLINE=89380219; PubMed=2528543;
RA Antonsson P., Heinegaard D., Oldberg A.;
RT "The keratan sulfate-enriched region of bovine cartilage proteoglycan
RT consists of a consecutively repeated hexapeptide motif.";
RL J. Biol. Chem. 264:16170-16173(1989).
RN [3]
RP SEQUENCE OF 1609-2113 AND 2151-2364 FROM N.A.
RX MEDLINE=87270630; PubMed=3111460;
RA Oldberg A., Antonsson P., Heinegaard D.;
RT "The partial amino acid sequence of bovine cartilage proteoglycan,
RT deduced from a cDNA clone, contains numerous Ser-Gly sequences
RT arranged in homologous repeats.";
RL Biochem. J. 243:255-259(1987).
RN [4]
RP SEQUENCE OF 2114-2150 FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93352525; PubMed=8349621;
RA Fuelleop C., Walcz E., Valyon M., Glant T.T.;
RT "Expression of alternatively spliced epidermal growth factor-like
RT domains in aggrecans of different species. Evidence for a novel
RT module.";
RL J. Biol. Chem. 268:17377-17383(1993).
RN [5]
RP PARTIAL SEQUENCE.
RX MEDLINE=85027710; PubMed=6489519;
RA Perin J.-P., Bonnet F., Jolles J., Jolles P.;
RT "Sequence data concerning the protein core of the cartilage
RT proteoglycan monomers. Characterization of a sequence allowing the
RT synthesis of an oligonucleotide probe.";
RL FEBS Lett. 176:37-42(1984).
RN [6]
RP PARTIAL SEQUENCE.
RX MEDLINE=87005253; PubMed=3530809;
RA Perin J.-P., Bonnet F., Jolles P.;
RT "Structural relationship between link proteins and proteoglycan
RT monomers";
RL FEBS Lett. 206:73-77(1986).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P13608-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P13608-2; Sequence=VSP_003072;
CC TERMINUS: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE C-TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND O-
CC LINKED (ABOUT 40) OLIGOSACCHARIDES.
CC -1- PTM: THE KERATAN SULFATE CONTENTS DIFFER CONSIDERABLY BETWEEN
CC ADULT AND FETAL BOVINE PROTEOGLYCANS.

CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 4 link domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U76615; AAB38524.1; -;
CC EMBL: L07053; -; NOT_ANNOTATED_CDS.
CC PIR: A34234; A39808.
CC PIR: T42630; T42630.
CC HSSP: P08709; 1BF9.
CC InterPro: IPR002353; AntifreezeII.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000742; EGF 2.
CC InterPro: IPR001881; EGF Ca.
CC InterPro: IPR006209; EGF-like.
CC InterPro: IPR007110; Ig-Like.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR001304; Lectin_C.
CC InterPro: IPR000538; Link.
CC InterPro: IPR003324; SGXSG.
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC Pfam: PF00008; EGF; 1.
CC Pfam: PF00047; Ig; 1.
CC Pfam: PF00059; lectin_c; 1.
CC Pfam: PF02339; SGXSG; 61.
CC Pfam: PF00084; sushi; 1.
CC Pfam: PF00193; Xlink; 4.
CC PRINTS: PR01265; LINKMODULE.
CC PRINTS: PR00356; ANTIFREEZEII.
CC ProDom: PD000918; Link; 4.
CC SMART: SM00032; CCP; 1.
CC SMART: SM00034; CLECT; 1.
CC SMART: SM00179; EGF CA; 1.
CC SMART: SM00445; LINK; 4.
CC PROSITE: PS00010; ASX_HYDROXYL; 1.
CC PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
CC PROSITE: PS00041; C_TYPE_LECTIN_2; 1.
CC PROSITE: PS00022; EGF_1; 1.
CC PROSITE: PS01187; EGF CA; 1.
CC PROSITE: PS00835; IG LIKE; 1.
CC PROSITE: PS00290; IG_MHC; FALSE_NEG.
CC PROSITE: PS01241; LINK; 4.
CC Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
CC EGF-like domain; Calcium; Alternative splicing; Repeat;
CC Immunoglobulin domain.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 2364 AGGREGAN CORE PROTEIN.
FT DOMAIN 25 147 IG-LIKE V-TYPE.
FT DOMAIN 170 247 LINK 1.
FT DOMAIN 268 349 LINK 2.
FT DOMAIN 504 581 LINK 3.
FT DOMAIN 602 683 LINK 4.
FT DOMAIN 774 907 23 X 6 AA APPROXIMATE TANDEM REPEATS OF
FT E-[EK]-P-P-P-S.
FT CS-2.
FT DOMAIN 1433 2112 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2113 2149 G3.
FT DOMAIN 2114 2364 C-TYPE LECTIN.
FT DOMAIN 2161 2276 SUSHI.
FT DOMAIN 2280 2338 BY SIMILARITY.
FT DISULFID 51 133 BY SIMILARITY.
FT DISULFID 175 246 BY SIMILARITY.
FT DISULFID 199 220 BY SIMILARITY.
FT DISULFID 273 348 BY SIMILARITY.

```

FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 509 580 BY SIMILARITY.
FT DISULFID 533 554 BY SIMILARITY.
FT DISULFID 607 682 BY SIMILARITY.
FT DISULFID 631 652 BY SIMILARITY.
FT DISULFID 2117 2128 BY SIMILARITY.
FT DISULFID 2182 2274 BY SIMILARITY.
FT DISULFID 2250 2266 BY SIMILARITY.
FT DISULFID 2281 2324 BY SIMILARITY.
FT DISULFID 2310 2337 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VAKSPDIC 2114 2150 Missing (in isoform 2).
FT VAKSPDIC 2114 2150 /FTID=VSP_003072.
SQ SEQUENCE 2364 AA; 246359 MW; 6FF83763420C3D4C CRC64;

Query Match 14.2k; Score 158.5; DB 1; Length 2364;
Best Local Similarity 26.8k; Pred. No. 7.8e-06;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;

Qy 9 QPVCRGG---TQRPCKYVYFHDTSRLNFEAKEACRRDGGQLVSVIESEDEQKLEKFI 65
Db 2151 QKCEEGWTKFQGHYR--HPFD---RATWVDASQCRKQKQSHLSIVTPEEQ-----EFV 2201

Qy 66 ENLLPSDQDP-WIGLRREEKQSNSTACQDIYATDGSISQFRNYYVDEP-----SCSEV 120
Db 2202 NN---NAQDYQWIGL-----NDKTIEGDFRMSDGHSLQFENWRPNQDNFPATGEDC 2250

Qy 121 CVVMYHOPSAPAGIGPGPMFOWNDRCNMKNKNEFKCKYS-----DEKPAVPSREARGEETE 175
Db 2251 VMIWHEKG-----EWNVPVNYQLPFTCKKGTVACGPPVVEHARIFGQKQD 2298

RESULT 14
PGCA HUMAN
ID PGCA HUMAN STANDARD; PRT: 2415 AA.
AC P16112; Q13650;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSPCP) (Chondroitin sulfate proteoglycan core protein 1).
GN AGC1 OR CSPG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Chondrocytes;
RX MEDLINE=91093289; PubMed=1985970;
RA Doegge K.J., Sasaki M., Kimura T., Yamada Y.;
RT "Complete coding sequence and deduced primary structure of the human
RT cartilage large aggregating proteoglycan, aggrecan. Human-specific
RT repeats, and additional alternatively spliced forms.";
RL J. Biol. Chem. 266:894-902(1991).
RN [2]
RP SEQUENCE OF 1778-2415 FROM N.A. (ISOFORM 2).
RC TISSUE=Chondrocytes;
RA Dudhia J., Hardingham T.E.;
RL Submitted (JAN-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1936-2415 FROM N.A. (ISOFORM 1).
RX MEDLINE=89380154; PubMed=2789216;
RA Baldwin C.T., Reginato A.M., Prockop D.J.;
RT "A new epidermal growth factor-like domain in the human core protein
RT for the large cartilage-specific proteoglycan. Evidence for
RT alternative splicing of the domain.";
RL J. Biol. Chem. 264:15747-15750(1989).

```

```

RN [4]
RP SEQUENCE OF 764-864 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=95128522; PubMed=7827755;
RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
RT "Length variation in the keratan sulfate domain of mammalian
RT aggrecan.";
RL Matrix Biol. 14:323-328(1994).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION.
CC -1- SUBUNIT: Interacts with FBLN1 (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P16112-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P16112-2; Sequence=VSP_003074;
CC Name=3;
CC IsoId=P16112-3; Sequence=VSP_003074, VSP_003075;
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, C MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 4 link domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M55172; AAA62824.1; -
CC EMBL; J05062; AAA35726.1; -
CC EMBL; X17405; CAA35463.1; -
CC EMBL; S74659; AAC60643.2; -
CC PIR; A39086; A39086.
CC HSSP; P98066; ITSG.
CC Genew; HGNC:319; AGC1.
CC MTM; 155760; -
CC GO; GO:0005204; F.chondroitin sulfate proteoglycan; TAS.
CC InterPro; IPR02353; Antifreeze1.
CC InterPro; IPR006209; EGF_like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR005338; Link.
CC InterPro; IPR003324; SGXXSG.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00047; ig; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF02339; SGXXSG; 71.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 4.

```

DR PRINTS; PRO1265; LINKMODULE.
 DR PRINTS; PRO0356; ANTIFREEZEII.
 DR ProDom; PD000918; Link; 4.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 4.
 DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN 2; 1.
 DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.
 DR PROSITE; PS00022; EGF 1; 1.
 DR PROSITE; PS01186; EGF 2; 1.
 DR PROSITE; PS00835; IG-LIKE; 1.
 DR PROSITE; PS00290; IG-MHC; 1.
 DR PROSITE; PS01241; LINK; 3.
 DR Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW EGF-like domain; Alernative splicing; Repeat; Immunoglobulin domain.
 FT SIGNAL 1 19
 FT CHAIN 20 2415
 FT DOMAIN 34 147
 FT DOMAIN 170 247
 FT DOMAIN 268 349
 FT DOMAIN 495 572
 FT DOMAIN 593 673
 FT DOMAIN 2164 2199
 FT DOMAIN 2201 2327
 FT DOMAIN 2331 2389
 FT DOMAIN 48 141
 FT DOMAIN 152 247
 FT DOMAIN 253 349
 FT DOMAIN 477 571
 FT DOMAIN 578 672
 FT DOMAIN 676 848
 FT DOMAIN 772 844
 FT DOMAIN 851 1497
 FT DOMAIN 941 1497
 FT DOMAIN 1498 2162
 FT DOMAIN 2163 2415
 FT DISULFID 51 133
 FT DISULFID 175 246
 FT DISULFID 199 220
 FT DISULFID 273 348
 FT DISULFID 297 318
 FT DISULFID 500 571
 FT DISULFID 524 545
 FT DISULFID 598 672
 FT DISULFID 621 642
 FT DISULFID 2168 2178
 FT DISULFID 2173 2187
 FT DISULFID 2189 2198
 FT DISULFID 2205 2216
 FT DISULFID 2233 2325
 FT DISULFID 2301 2317
 FT DISULFID 2332 2375
 FT DISULFID 2361 2388
 FT CARBOHYD 126 126
 FT CARBOHYD 239 239
 FT CARBOHYD 333 333
 FT CARBOHYD 387 387
 FT CARBOHYD 434 434
 FT CARBOHYD 602 602
 FT CARBOHYD 657 657
 FT CARBOHYD 737 737
 FT CARBOHYD 1898 1898
 FT VARSPLIC 2163 2200
 FT VARSPLIC 2330 2390
 FT CONFLICT 766 766
 FT CONFLICT 847 847
 FT CONFLICT 1928 1928
 FT CONFLICT 1964 1964
 FT CONFLICT 2070 2070

FT CONFLICT 2391 2391 A -> P (IN REF. 2 AND 3).
 SQ SEQUENCE 2415 AA; 250191 MW; 128893751B98C6B6 CRC64;
 Query Match 14.2%; Score 158.5; DB 1; Length 2415;
 Best Local Similarity 29.1%; Pred. No. 8.1e-06;
 Matches 52; Conservative 22; Mismatches 62; Indels 43; Gaps 11;
 QY 9 QPVCRRG--TORPCYKVIYFHDTSRLNFRFEAKACRRDGGQLVYSIESEDKLIEKFI 65
 DB 2202 QEVCEGWKYGHCYR--HFPD---RETWDAERRCRQSHLSIVTPEEQ-----EFV 2252
 QY 66 ENLLPSDGDFF-WIGLRREKQSNSTACQDLYAMTDGSIQFRNMYVDPS---CGSEVC 121
 DB 2253 NN---NAQDYQWIGL-----NDRTIEGDFRMSDGHMPQFENWRPNQDNFFAAGDC 2301
 QY 122 VVM-YHQPSAPAGIGPPYMFOWNDRCMKNKFNFKYS----DEKPAVPSREAGEETE 175
 DB 2302 VVMIWHEKG-----EMNDVPCNYHLPTCKGTGACGEPVVEHARTFOOKKD 2349
 RESULT 15
 PGCA_CANPA STANDARD; PRT; 2333 AA.
 ID PGCA_CANPA Q28310;
 AC Q28343; Q28310;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
 GN AGC1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Glant T.T., Adams M.E., Kwok S.X.F., Huang D., Fuloop C.;
 RT "Complete coding sequence and deduced amino acid sequence of aggrecan of canine cartilage."
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 774-833 FROM N.A.
 RC TISSUE=Cartilage;
 RX MEDLINE=9512852; PubMed=7827755;
 RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
 RT "Length variation in the keratan sulfate domain of mammalian aggrecan."
 RL Matrix Biol. 14:323-328(1994).
 RN [3]
 RP SEQUENCE OF 1830-2333 FROM N.A.
 RA Adams M.E., Kwok S.X.F., Huang D., Glant T.T., Fuloop C.;
 RT Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 2082-2118 FROM N.A.
 RC TISSUE=Cartilage;
 RX MEDLINE=93352525; PubMed=8349621;
 RA Fuloop C., Walcz E., Valyon M., Glant T.T.;
 RT "Expression of alternatively spliced epidermal growth factor-like domains in aggrecans of different species. Evidence for a novel module."
 RL J. Biol. Chem. 268:17377-17383(1993).
 CC -I- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -I- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
 CC -I- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)

CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
 CC AND G3.
 CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
 CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -1- SIMILARITY: Contains 4 link domains.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -1- SIMILARITY: Contains 1 EGF-like domain.
 CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VESICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U65989; AAC06238.2; -;
 CC EMBL; S74662; AAC60527.1; -;
 CC PIR; L07054; -; NOT_ANNOTATED_CDS.
 CC HSP; P08709; 1BF9.
 CC InterPro; IPR000152; Aex_hydroxyl.
 CC InterPro; IPR000742; EGF_2.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF_like.
 CC InterPro; IPR007110; Ig_Like.
 CC InterPro; IPR003006; Ig_MHC.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR003038; Link.
 CC InterPro; IPR003324; SCXXSG.
 CC InterPro; IPR000436; Sushi_SCR_CCP.
 CC Pfam; PF00008; EGF; 1.
 CC Pfam; PF00047; Ig; 1.
 CC Pfam; PF00059; lectin_c; 1.
 CC Pfam; PF02339; SGXXSG; 66.
 CC Pfam; PF00084; sushi; 1.
 CC Pfam; PF00193; Xlink; 4.
 CC PRINTS; PR01265; LINKMODULE.
 CC PROSITE; PS00356; ANTIFREEZEII.
 CC ProDom; PD000918; Link; 4.
 CC SMART; SM00034; CLECT; 1.
 CC SMART; SM00179; EGF_CA; 1.
 CC SMART; SM00445; LINK; 4.
 CC PROSITE; PS00010; ASX_HYDROXYL; 1.
 CC PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
 CC PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
 CC PROSITE; PS00022; EGF_1; 1.
 CC PROSITE; PS01187; EGF_CA; 1.
 CC PROSITE; PS00835; IG_LIKE; 1.
 CC PROSITE; PS00290; IG_MHC; 1.
 CC PROSITE; PS01241; LINK; 4.
 CC Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 CC EGF-like domain; Repeat; Immunoglobulin domain.
 CC SIGNAL 1 15 POTENTIAL
 CC CHAIN 17 2333 AGGRECAN CORE PROTEIN.
 CC DOMAIN 34 147 IG-LIKE V-TYPE.
 CC DOMAIN 170 247 LINK 1.
 CC DOMAIN 268 349 LINK 2.
 CC DOMAIN 513 590 LINK 3.
 CC DOMAIN 611 692 LINK 4.
 CC DOMAIN 2081 2117 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
 CC DOMAIN 2130 2245 C-TYPE LECTIN.
 CC DOMAIN 2249 2307 SUSHI.
 CC DOMAIN 48 140 G1-A.
 CC DOMAIN 152 247 G1-B.
 CC DOMAIN 253 349 G2-B.
 CC DOMAIN 495 589 G2-B.
 CC DOMAIN 596 691 G2-B.
 CC DOMAIN 694 816 KS.

| | | | | |
|----|----------|------|------------|-----------------------------------|
| FT | DOMAIN | 819 | 1394 | CS-1. |
| FT | DOMAIN | 1395 | 2079 | CS-2. |
| FT | DOMAIN | 2080 | 2333 | G3. |
| FT | DISULFID | 51 | 133 | BY SIMILARITY. |
| FT | DISULFID | 175 | 246 | BY SIMILARITY. |
| FT | DISULFID | 199 | 220 | BY SIMILARITY. |
| FT | DISULFID | 273 | 348 | BY SIMILARITY. |
| FT | DISULFID | 297 | 318 | BY SIMILARITY. |
| FT | DISULFID | 518 | 589 | BY SIMILARITY. |
| FT | DISULFID | 542 | 563 | BY SIMILARITY. |
| FT | DISULFID | 616 | 691 | BY SIMILARITY. |
| FT | DISULFID | 640 | 661 | BY SIMILARITY. |
| FT | DISULFID | 2085 | 2096 | BY SIMILARITY. |
| FT | DISULFID | 2090 | 2105 | BY SIMILARITY. |
| FT | DISULFID | 2107 | 2116 | BY SIMILARITY. |
| FT | DISULFID | 2123 | 2134 | BY SIMILARITY. |
| FT | DISULFID | 2151 | 2243 | BY SIMILARITY. |
| FT | DISULFID | 2219 | 2235 | BY SIMILARITY. |
| FT | DISULFID | 2250 | 2293 | BY SIMILARITY. |
| FT | DISULFID | 2279 | 2306 | BY SIMILARITY. |
| FT | CARBOHYD | 126 | 126 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 239 | 239 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 333 | 333 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 387 | 387 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 444 | 444 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 620 | 620 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 676 | 676 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| FT | CARBOHYD | 747 | 747 | N-LINKED (GLCNAC. .) (POTENTIAL). |
| SQ | SEQUENCE | 2333 | AA; 240573 | MW; 899ED78F3508B596 |

Query Match 13.9%; Score 155.5; DB 1; Length 2333;
 Best Local Similarity 28.5%; Pred. No. 1.4e-05;
 Matches 51; Conservative 24; Mismatches 61; Indels 43; Gaps 11;

| | | | | |
|----|------|---|--|----------|
| Qy | 9 | QVCRGG--- | TQPCYKVIYFHDTSRLNFEAEACRRDGGQLVSISEDEQKLIKFI | 65 |
| Db | 2120 | QELCEGWTKFQGHYR--YFPD--- | RESWVDAESRCRAQQSHLSIVTPEQ--- | EFV 2170 |
| Qy | 66 | ENLLPSDGF-WIGLRREKQSNSTACQDIYAWTQSGISQFRNYYVDEPS--- | CGSEVC | 121 |
| Db | 2171 | NN---NAQDYQWIGL----- | NDRTIEGDFRMSDGHSLQFENWRPNQPDNFFVSGDC | 2219 |
| Qy | 122 | VVM-YHOPSAPAGIGGYPFMQWDDRCNMKNPKICKYS--- | DEKPAVPSREASGEETE | 175 |
| Db | 2220 | VVMWHEKG----- | EWNDVPCNYLPFTCKKGTACGDPVVEHARTFGQKDD | 2267 |

Search completed: December 22, 2003, 16:10:57

Job time : 6.56161 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:06:34 ; Search time 10.0151 Seconds
(without alignments)
1978.090 Million cell updates/sec

Title: US-09-887-855-2_COPY_22_227

Perfect score: 1115
Sequence: 1 ATGRLLSGQPVCRGGTORPC.....EEDAKTKFSREAAALNLAY 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76.*

1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 182 | 16.3 | 1456 | 1 A36563 | mannose receptor p |
| 2 | 178.5 | 16.0 | 1455 | 1 A48925 | mannose receptor p |
| 3 | 177 | 15.9 | 1268 | 2 S52781 | neurocan - mouse |
| 4 | 174.5 | 15.7 | 1643 | 2 T14274 | versican precursor |
| 5 | 174.5 | 15.7 | 3381 | 2 T42389 | versican precursor |
| 6 | 174 | 15.6 | 1257 | 2 S28764 | neurocan precursor |
| 7 | 174 | 15.6 | 2397 | 1 A55535 | versican precursor |
| 8 | 174 | 15.6 | 2409 | 1 A60979 | versican precursor |
| 9 | 171 | 15.3 | 3562 | 2 A47171 | chondroitin sulfat |
| 10 | 170.5 | 15.3 | 1479 | 2 T42710 | mannose receptor, |
| 11 | 158.5 | 14.2 | 1340 | 2 A39808 | proteoglycan core |
| 12 | 158.5 | 14.2 | 2327 | 2 T42630 | aggreccan - bovine |
| 13 | 158.5 | 14.2 | 2415 | 1 A39086 | aggreccan precursor |
| 14 | 154.5 | 13.9 | 612 | 2 B42755 | E-selectin precurs |
| 15 | 153.5 | 13.8 | 2124 | 2 A28452 | proteoglycan core |
| 16 | 152 | 13.6 | 912 | 2 A54423 | brevican precursor |
| 17 | 149.5 | 13.4 | 459 | 2 T24425 | hypothetical prote |
| 18 | 149 | 13.4 | 321 | 1 LNHUP1 | IGF Fc receptor II |
| 19 | 148.5 | 13.3 | 330 | 2 T46256 | brevican - human (|
| 20 | 148.5 | 13.3 | 2132 | 1 A55182 | aggreccan precursor |
| 21 | 148 | 13.3 | 253 | 2 E89130 | protein F52E1.2 [i |
| 22 | 147 | 13.2 | 893 | 2 S57653 | brevican precursor |
| 23 | 146.5 | 13.1 | 162 | 1 LNRCL1 | lectin BRA3-1 prec |
| 24 | 146 | 13.1 | 2109 | 1 I50421 | aggreccan precursor |
| 25 | 145.5 | 13.0 | 742 | 2 JC7595 | scavenger receptor |
| 26 | 145 | 13.0 | 883 | 2 S49126 | brevican precursor |
| 27 | 144.5 | 13.0 | 173 | 2 S10548 | lectin - barnacle |
| 28 | 144.5 | 13.0 | 372 | 2 S23936 | L-selectin precurs |
| 29 | 144.5 | 13.0 | 404 | 2 A46274 | HIV gp120-binding |

| | | | | | |
|----|-------|------|------|----------|--------------------|
| 30 | 143.5 | 12.9 | 129 | 2 JC4329 | coagulation factor |
| 31 | 143.5 | 12.9 | 372 | 1 A32375 | L-selectin precurs |
| 32 | 143.5 | 12.9 | 463 | 2 T26655 | hypothetical prote |
| 33 | 142.5 | 12.8 | 131 | 2 JCS058 | bisectin alpha c |
| 34 | 142 | 12.7 | 1487 | 2 S48719 | phospholipase-A(2) |
| 35 | 141.5 | 12.7 | 331 | 1 LNM5ER | IGF Fc receptor, I |
| 36 | 140.5 | 12.6 | 162 | 1 LNRCS3 | lectin BRA3-2 prec |
| 37 | 140 | 12.6 | 370 | 2 S22124 | L-selectin precurs |
| 38 | 139 | 12.5 | 248 | 1 LNHUPS | pulmonary surfacta |
| 39 | 139 | 12.5 | 248 | 1 LNHUP6 | pulmonary surfacta |
| 40 | 139 | 12.5 | 248 | 1 LNHUP1 | pulmonary surfacta |
| 41 | 139 | 12.5 | 283 | 1 LNFHLS | lectin precursor - |
| 42 | 138.5 | 12.4 | 152 | 2 JC4690 | coagulation factor |
| 43 | 138.5 | 12.4 | 202 | 2 JC4031 | tetranectin precur |
| 44 | 138 | 12.4 | 280 | 2 T29200 | hypothetical prote |
| 45 | 137.5 | 12.3 | 309 | 1 S34198 | IGF Fc receptor II |

ALIGNMENTS

RESULT 1

A36563

mannose receptor precursor - human

C:Species: Homo sapiens (man)

C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999

C:Accession: A36563; A60926; A44255; B44255; C44255; D44255; E44255; F44255; G44255; H:

R:Taylor, M.E.; Conary, J.T.; Lennartz, M.R.; Stahl, P.D.; Drickamer, K.

J. Biol. Chem. 265, 12156-12162, 1990

A>Title: Primary structure of the mannose receptor contains multiple motifs resembling

A:Reference number: A36563; MUID:90324192; PMID:2373685

A:Accession: A36563

A:Molecule type: mRNA

A:Residues: 1-1456 <TAY>

A:Cross-references: GB:J05550; NID:9188675; PIDN:AAA59868.1; PID:9188676

A>Note: parts of this sequence, including the amino end of the mature protein, were coi

R:Exekowitz, R.A.B.; Sastry, K.; Bailly, P.; Warner, A.

J. Exp. Med. 172, 1785-1794, 1990

A>Title: Molecular characterization of the human macrophage mannose receptor: demonstr

A:Reference number: A60926; MUID:91079783; PMID:2258707

A:Accession: A60926

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1333,'T',1335-1456 <EZE>

A:Cross-references: GB:X55635

A>Note: translation of the nucleotide sequence is incomplete

R:Kim, S.J.; Ruiz, N.; Bezouska, K.; Drickamer, K.

Genomics 14, 721-727, 1992

A>Title: Organization of the gene encoding the human macrophage mannose receptor (MRC1

A:Reference number: A44255; MUID:93052405; PMID:1294118

A:Accession: A44255

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual t:

A:Molecule type: DNA

A:Residues: 155-233,'KSAL',238-283;346-428;492-569;631-714,716-719;783-820,'N',822-865

A>Note: sequence extracted from NCBI backbone (NCBIP:118415, NCBIP:118421, NCBIP:118442;

C:Genetics:

A:Gene: GDB:MRC1

A:Cross-references: GDB:133759; OMIM:153618

A:Map position: 10p13-10p13

C:Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II :

C:Keywords: duplication; lectin; tandem repeat; transmembrane protein

F:1-18/Domain: signal sequence #status predicted <SIG>

F:168-209/Domain: fibronectin type II repeat homology <2F1>

F:223-340/Domain: C-type lectin homology <LCH1>

F:362-486/Domain: C-type lectin homology <LCH2>

F:945-1079/Domain: C-type lectin homology <LCH3>

Query Match 16.3%; Score 182; DB 1; Length 1456;

Best Local Similarity 25.5%; Pred. No. 9.1e-08;

Matches 52; Conservative 37; Mismatches 61; Indels 54; Gaps 8;

QY 21 YKVIYFHTDTSRLNFEAKEACRRDGGQLVSISEDEQKLIENLLPSDGFWIGLR 80

```
Db      807 YKDYQYFSEKETMDNARAFCKNFGDLVSIQSESEKFLWKYV-NRNDAQSAYFIGLL 865
QY      81 RREEKQSNSTACODLYAWTDGSIQFRNWTYVDPSCGS--EVCVMYHOPSPAGIGPY 138
Db      866 ISLDKK-----FAWMDGSKVDYVSMATGEPNPFANEDENCVTWY-----SNSGF---- 908
QY      139 MFWNDDRCNMKNFICKYSDEK----PAVPSREASGEETELTPVLPEETQE----- 187
Db      909 ----WNIDNGYPNFIQRRHNSINATVMP-----TMPSPGCKEGMNFYSN 954
QY      188 -----EDAKTKFKESREAAAL 202
Db      955 KCFKIFGFMEERKNQWAKACI 978

RESULT 2
A48925
mannose receptor precursor, macrophage - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A48925; S21320; PC2245
R:Harris, N.; Super, M.; Rits, M.; Chang, G.; Ezekowitz, R.A.
Blood 80, 2363-2373, 1992
A:Title: Characterization of the murine macrophage mannose receptor: demonstration that
A:Reference number: A48925; MUID:93043353; PMID:1421407
A:Accession: A48925
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1455 <HA2>
A:Experimental source: peritoneal macrophage
A>Note: sequence extracted from NCBI backbone (NCBIP:118733)
R:Harris, N.; Super, M.; Rits, M.; Chang, G.; Ezekowitz, R.B.
submitted to the EMBL Data Library, April 1992
A:Description: Characterization of the murine macrophage mannose receptor: Demonstration
on.
A:Reference number: S21320
A:Accession: S21320
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-302, 'W', 303-1117, 'E', 1119-1455 <HA2>
A:Cross-references: EMBL:Z11974; NID:G52997; PIDN:CAA78028.1; PID:G52998
R:Harris, N.; Peters, L.L.; Eicher, E.M.; Rits, M.; Raspberry, D.; Eichbaum, Q.G.; Super
Biochem. Biophys. Res. Commun. 198, 682-692, 1994
A:Title: The exon-intron structure and chromosomal localization of the mouse macrophage
A:Reference number: PC2245; MUID:94128116; PMID:8297379
A:Accession: PC2245
A:Molecule type: mRNA
A:Residues: 35-105 <HA3>
C:Genetics:
A:Gene: Mrcl
A:Map position: 2
C:Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II re
C:Keywords: membrane protein; receptor
F:168-209/Domain: fibronectin type II repeat homology <2F9>
F:361-485/Domain: C-type lectin homology <LCHI>
F:943-1077/Domain: C-type lectin homology <LCH2>

Query Match 16.0%; Score 178.5; DB 1; Length 1455;
Best Local Similarity 25.9%; Pred. No. 1.8e-07;
Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;

QY      21 YKVIYFHDTSRLNFEAKEACRRDGGQLVSIQSEDEQKLIKFIENLLPSDGFWIGLR 80
Db      806 YKDYQYFSEKETMDNARFCKNFGDLATIKSESEKFLWKYI-NKNGGSPYFIGML 864
QY      81 RREEKQSNSTACODLYAWTDGSIQFRNWTYVDPSCGS--EVCVMYHOPSPAGIGPY 138
Db      865 ISMDKK-----FIWMDGSKVDVFAWATGEPNPFANDENCVTWY-----TNSGF---- 907
QY      139 MFWNDDRCNMKNFICK--YSDEKPAVPSREASGEETELTPVLPEETQE----- 187
Db      908 ----WNIDNGYPNFIQRRHNSINATAMP-----TPTTPGCKEGMHLKYNK 953
```

```
QY      188 -----EDAKTKFKESREAAALNL 204
Db      954 CPKIFGFANBEKKSWQDARQACKGL 978

RESULT 3
S52781
neurocan - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 04-Feb-2000
C:Accession: S52781
R:Rauch, U.; Foreberg, N.; Kulbe, G.; Arnold-Ammer, I.; Faessler, R.
submitted to the EMBL Data Library, February 1995
A:Description: Amino acid sequence of mouse neurocan and brevican and their different e
A:Reference number: S52781
A:Accession: S52781
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1268 <RAU>
A:Cross-references: EMBL:X84727; NID:G758629; PIDN:CAA59216.1; PID:G758630
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; E
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-355/Domain: link protein repeat homology <LNK2>
F:964-995/Domain: EGF homology <EGF>
F:1040-1160/Domain: EGF homology <EGF>
F:1167-1223/Domain: C-type lectin homology <LCH>
F:1167-1223/Domain: complement factor H repeat homology <PHD>

Query Match 15.9%; Score 177; DB 2; Length 1268;
Best Local Similarity 31.5%; Pred. No. 2.1e-07;
Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;

QY      17 QRPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSIQSEDEQKLIKFIENLLPSDGF 76
Db      1048 QGCHYR--YF---AHRRAWEDAECDRCRRAGHLTSVHSPEEHKFINSF-----GHNSW 1096

QY      77 IGLRREEKQSNSTACQDLYAWTDGSIQFRNWTYVDPERS--CGSEVCVMYHQPSPAG 133
Db      1097 IGLNDRTVRD-----FQWTDNTGLQYENRREKQPNFFAGGEDCVMVAHESG--- 1145

QY      134 IGGPYMFQWDDRCNMKNFICK 156
Db      1146 -----RWNDVPCYNLPYVCK 1161

RESULT 4
T14274
versican precursor, splice form V2 - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000
C:Accession: T14274
R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
J. Biol. Chem. 273, 15758-15764, 1998
A:Title: Versican V2 is a major extracellular matrix component of the mature bovine bra
A:Reference number: Z17954; MUID:98288320; PMID:9624174
A:Accession: T14274
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1643 <SCH>
A:Cross-references: EMBL:AF060458; NID:G3253303; PID:G3253304; PIDN:AAC24360.1
A:Experimental source: brain
C:Keywords: glycoprotein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-1643/Product: versican, splice form V2 #status predicted <MAT>
F:57,331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (Asn) (covalent) #s

Query Match 15.7%; Score 174.5; DB 2; Length 1643;
Best Local Similarity 25.5%; Pred. No. 4.8e-07;
Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;

QY      17 QRPCYKVIYFHDTSRLNFEAKEACRRDGGQLVSIQSEDEQKLIKFIENLLPSDGF 75
Db      1424 QGCHYR--YF---AHRRTWDAARECRQGAHLTSILSHEEQMFINRV-----GHDYQ 1471
```


| | | | | |
|----|------|--|-------------------|-----|
| Qy | 76 | WIGLRREEKQSNSTACODLYAMTDGSGISQFRNWWYDEP---- | SCGSEVCVWYWHQPSAP | 131 |
| | | | | |
| Db | 1472 | WIGL-----NDKMFEDFRWTDGSLTLOYENRPNQDPSFFSTGDCVWIWHENG-- | 1521 | |
| | | | | |
| Qy | 132 | AGIGGYMFOWNDDRCMMKNFNICKYS-----DEKPAVPSREAEGE----- | 172 | |
| | | | | |
| Db | 1522 | -----QWNDVPCNYHLTYTCKTKVACGPPVVENAKTFGKMKPRYEINSLIRYHC | 1572 | |
| | | | | |
| Qy | 173 | -----ETELTT-----PVL-----PEETQEEADAKKTFKESRAALN | 203 | |
| | | | | |
| Db | 1573 | KDGFQIRHLLPTIRCLNGRWAMPKAITCLNPSAYQRTYSKKYFNQSSSAKDN | 1623 | |
| | | | | |

RESULT 5
T42389
vesicarin precursor, splice form V0 - bovine
N;Alternate names: chondroitin sulfate proteoglycan
C;Species: Bos primigenius taurus (cattle)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 05-May-2000
C;Accession: T42389
R;Schmalfeldt, M.; Doure-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
J. Biol. Chem. 273, 15758-15764, 1998
A;Title: Vesican V2 is a major extracellular matrix component of the mature bovine brain
A;Reference number: Z17954; MUID:98288320; PMID:9624174
A;Accession: T42389
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1-3381 <SCH>
A;Cross-references: EMBL:AF060456; NID:G3253299; PID:G3253300; PIDN:AAC24358.1
C;Superfamily: chicken chondroitin sulfate proteoglycan Pg-M core protein; C-type lectin
C;Keywords: chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein
F;1-domain: signal sequence #status predicted <SIG>
F;21-3381/Product: vesicarin, splice form V0 #status predicted <MAT>
F;57,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2356,2382
F;57,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2356,2382

| | | | | | |
|-----------------------|-------|--|---------------|-------------|---------|
| Query Match | 15.7% | Score 174.5 | DB 2 | Length 3381 | |
| Best Local Similarity | 25.5% | Pred. No. 1.1e-06 | | | |
| Matches | 59 | Conservative 26 | Mismatches 71 | Indels 75 | Gaps 11 |
| QY | 17 | QRPCYKVIYPHDTSRRLNPFEEAKRCRDGGQLVSTESDEQKLIKFIENLLPSSDGF- | 75 | | |
| Db | 3162 | QGQCYK--YP--AHRRTWDAARECRLQGAHLTSLSHSEQMFVNRV-----GHDIQ | 3209 | | |
| QY | 76 | WIGLRREEKQSNSTACODLIYATDGSISQFRWYVDEP-----SCSEVCVMVWHPSPAP | 131 | | |
| Db | 3210 | WIGL-----NDKMFEDHFRWTDGSTLOYENRRNPQDPSFSTGDCDVIWHENG-- | 3259 | | |
| QY | 132 | AGIGGPMYFOWNDDRCNMKGNFTCKYS-----DKPAPVPSRAGE----- | 172 | | |
| Db | 3260 | -----QWNDVPCNYHLTYTCKGTACGPPVVENAKTGGKMPRYEINSLTRYHC | 3310 | | |
| QY | 173 | -----ETELTT-----PVL-----PEETQEEADAKTKPKESREALN | 203 | | |
| Db | 3311 | KQGFIRHLPTIRCLNGRWAMKPTICLNPSAYQRTYSKKYFNKSSAKON | 3361 | | |

RESULT 6
S28764
neurocan precursor - rat
S28764
C:Species: Rattus norvegicus (Norway rat)
C>Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
C:Accession: S28764
R:Rauch, U.; Karthikeyan, L.; Maurel, P.; Margolis, R.K.
J. Biol. Chem. 267, 19536-19547, 1992
A:Title: Cloning and primary structure of neurocan, a developmentally regulated, aggregated proteoglycan core protein from rat brain
A:Reference number: S28764; MUID:92406907; PMID:1326557
A:Accession: S28764
A:Molecule type: mRNA
A:Residues: 1-1257 <RAU>
A:Cross-references: EMBL:M97161; NID:G205649; PIDN:AAC37679.1; PID:G205650
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EGF repeat homology
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein

F/1-22/Domain: signal sequence #status predicted <SIG>
F/23-1357/Product: neurocan #status predicted <NAT>
F/176-353/Domain: link protein repeat homology <LNKL1>
F/274-353/Domain: link protein repeat homology <LNK2>
F/364-366/Region: cell attachment (R-G-D) motif
F/953-984/Domain: EGF homology <EGF>
F/1029-1149/Domain: C-type lectin homology <LCH>
F/1156-1212/Domain: complement factor H repeat homology <FHD>
F/121-339,737,967,1164/Binding site: carbohydrate (Asn) (covalent) #status predicted
F/372,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F/944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

| | | | | |
|-----------------------|--------------|--------------------|----------------|--------------|
| Query Match | 15.6% | Score 174; | DB 2; | Length 1257; |
| Best Local Similarity | 30.8% | Pred. No. 3.9e-07; | | |
| Matches 44; | Conservative | 18; | Mismatches 49; | Indels 32; |
| | | | Gaps 6; | |

| | | | | |
|----|------|--|--------------------------|------|
| Qy | 17 | QRPCYKVIYPHDTSRRLNPEEAKRCRRDGGQLVSI | ESEDEQKLISKFIENLLPSDGDFW | 76 |
| Dz | : | : | : | : |
| Db | 1037 | QGHCYR--YF---AHRRAWEADERCRRAGHLTSVHS | PPEHKFINSF-----GHNSW | 1085 |
| Qy | 77 | IGLRRREEKQSNTACODLYATGDISIFRNMYVDPS--- | CGSEVCVVMMYHQPSAPAG | 133 |
| Dz | : | : | : | : |
| Db | 1086 | IGLNDRTVRD-----FQMTDNTGLQVENREKQPONF | FAGGEDCVMVAHENG--- | 1134 |
| Qy | 134 | IGGPYMFWQNDDBCNKNPFICK | 156 | |
| Dz | : | : | : | : |
| Db | 1135 | -----RWNDVPCYNLPYVCK | 1150 | |

RESULT 7
A55535
versican precursor - mouse
N;Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan
versican
N;Contains: glial hyaluronate-binding protein
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A55535
R;Ito, K.; Shinomura, T.; Zako, M.; Ujita, M.; Kimata, K.
J. Biol. Chem. 270, 958-965, 1995
A;Title: Multiple forms of mouse PG-M, a large chondroitin sulfate proteoglycan general
A;Reference number: A55535; MUID:95122551; PMID:7822336
A;Accession: A55535

A;Accession: A35353
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1-2397 <RES>
A;Cross-references: GB:D16263; NID:G862460; PIDN:BAA03796.1; PID:G862461
C;Superfamily: versican; C-type lectin homology; complement factor H repeat homology; I
F;1-20/Domain: signal sequence #status predicted <SIG>
F;161-1654/Domain: versican #status predicted <NAT>
F;167-244/Domain: link protein repeat homology <LNK1>
P;265-346/Domain: link protein repeat homology <LNK2>
P;2095-2126/Domain: EGF homology <EG1>
F;2133-2184/Domain: EGF homology <EG2>
F;2171-2291/Domain: C-type lectin homology <LCH>
F;2298-2354/Domain: complement factor H repeat homology <PHD>

| | | | | |
|-----------------------|------------------|--------------------|------------|--------------|
| Query Match | 15.6% | Score 174; | DB 1; | Length 2397; |
| Best Local Similarity | 28.5%; | Pred. NO. 8.3e-07; | | |
| Matches 47; | Conservative 23; | Mismatches 55; | Indels 40; | Gaps 8; |

| | | | |
|----|------|---|------|
| Qy | 17 | ORPCYKVIYFHDTSRLNFEFEAKACRRDGGQGVSTESDEOKLIEKPTENLLPSDGF- | 75 |
| Db | 2179 | QCQCYK--YF---AHRRTWDAERECRLQGAHLTSLSHEEQMFNVRV-----GHDVQ | 2226 |
| Qy | 76 | WIGLRRREEKQSNSTACODLYAWTDGSGISQFRNYYVDEP---SCGSEYCVVMYHQSAP | 131 |
| Db | 2227 | WIGL-----NDKNFEHDFRMTDGSALQYENWRPNQDPSFPSAGEDCVIIWHENG-- | 2276 |
| Qy | 132 | AGIGGYMFQWNRDRCNMKNPFICKYS-----DEKPAVFSRBAEGE | 172 |
| Db | 2277 | -----QWMDVFCNTHLYTTCCKGTGTVACQCPVVENAKTFGK | 2312 |

F;265-346/Domain: link protein repeat homology <LNK2>
F;559-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>
F;2106-2137/Domain: EGF homology <EG1>
F;2144-2175/Domain: EGF homology <EG2>
F;2182-2302/Domain: C-type lectin homology <LCH>
F;2309-2365/Domain: complement factor H repeat homology <FHD>

Query Match 15.6%; Score 174; DB 1; Length 2409;
Best Local Similarity 28.5%; Pred. No. 8.3e-07;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8

Qy 17 QRPCYKVIYPHDTSRRLNPFEEAKCRDGGQLVSIESDEQKLIEKFLENLLPSDGDFF- 75
| | | | | : : : : :
Db 2190 QGQCYK--YF--AHRTWDAARECRLQGAHLTSLSHSEEQMFVNRV-----GHDIQ 2237
| | | | | : : : : :
Qy 76 WIGLRREKQSINSTACQDLIYATDGISIQFRWYVDPE----SCGEVCVMVWHQSPAP 131
| | | | | : : : : :
Db 2238 WIGL-----NDKMFERDFRWTGSTGLQYENWRNPQDPSPFSAGEDCVIIWHENG-- 2287
| | | | | : : : : :
Qy 132 AGIGGPYMFQWDDRCNMKNFNFKYS-----DEKPVPSPREAAGE 172
| | | | | : : : : :
Db 2288 -----QWNVPCNYHLTYTCCKGTVACGPPPVVENAKTPGK 2323
| | | | | : : : : :

RESULT 9
A47171
chondroitin sulfate proteoglycan PG-M core protein - chicken
C;Species: Gallus gallus (chicken)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C;Accession: A47171
R;Shinomura, T.; Nishida, Y.; Ito, K.; Kimata, K.
J. Biol. Chem. 268, 14461-14469, 1993
A;Title: cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan expressed in chick embryo
A;Reference number: A47171; PMID:93300846; PMID:8314802
A;Accession: A47171
A>Status: preliminary
A:Molecule type: nucleic acid
A;Residues: 1-3562 <SH1>
A;Cross-references: GB:D13542; NID:G391643; PIDN:BAA02742.1; PID:G391644
A;Experimental source: stage 22-23 developing limb buds
A;Note: sequence extracted from NCBI backbone (NCBIN:L34456, NCBIP:L34457)
C;Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lectin
F;166-243/Domain: link protein repeat homology <LNK1>
F;264-345/Domain: link protein repeat homology <LNK2>
F;3258-3289/Domain: EGF homology <EGP>
F;3296-3327/Domain: EGF homology <EGP>
F;3334-3454/Domain: C-type lectin homology <LCH>
F;3461-3517/Domain: complement factor H repeat homology <FHD>

Query Match 15.3%; Score 171; DB 2; Length 3562;
Best Local Similarity 28.5%; Pred. No. 2.4e-06;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8

Qy 17 QRPCYKVIYPHDTSRRLNPFEEAKCRDGGQLVSIESDEQKLIEKFLENLLPSDGDFF- 75
| | | | | : : : : :
Db 3342 QGQCYK--YF--AHRTWDAARECRLQGAHLTSLSHSEEQMFVNRI-----GHDIQ 3389
| | | | | : : : : :
Qy 76 WIGLRREKQSINSTACQDLIYATDGISIQFRWYVDPE----SCGEVCVMVWHQSPAP 131
| | | | | : : : : :
Db 3390 WIGL-----NDKMFERDFRWTGSTGLQYENWRNPQDPSPFSAGEDCVIIWHENG-- 3439
| | | | | : : : : :
Qy 132 AGIGGPYMFQWDDRCNMKNFNFKYS-----DEKPVPSPREAAGE 172
| | | | | : : : : :
Db 3440 -----QWNVPCNYHLTYTCCKGTVACGPPPVVENAKTPGK 3475
| | | | | : : : : :

RESULT 10
T42710
mannose receptor, macrophage - mouse
N;Alternate names: lambda lectin; phospholipase A2 receptor
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000

A:Cross-references: GB:J05062; NID:g181167; PIDN:AAA35726.1; PID:g181168

C:Genetics:

A:Gene: GDB:AGCL1; CSFG1; CSPGCP; MSK16

A:Cross-references: GDB:127479; OMIM:155760

A:Map position: 15q26-15q26

C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; E

C:Keywords: alternative splicing; cartilage; chondroitin sulfate proteoglycan; extracel

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-2415/Product: aggrecan cartilage long splice form #status predicted <MATL>

F:20-2162,2201-2415/Product: aggrecan cartilage short splice form #status predicted <MA

F:20-2162,2201-2329,'A','2392-2415/Product: aggrecan short splice form #status predicted

F:44-135/Domain: immunoglobulin homology <IWM>

F:170-247/Domain: link protein repeat homology <LNK1>

F:248-349/Domain: link protein repeat homology <LNK2>

F:495-572/Domain: link protein repeat homology <LNK3>

F:697-861/Domain: keratan sulfate attachment #status predicted <KSA>

F:864-1510/Domain: chondroitin sulfate attachment #status predicted <CS1>

F:1511-2162/Domain: chondroitin sulfate attachment #status predicted <CS2>

F:2168-2198/Domain: EGF homology <EGF>

F:2305-2325/Domain: C-type lectin homology <LCH>

F:2332-2388/Domain: Complement factor H repeat homology <PHD>

F:126,239,333,387,434,602,657,737,1898/Binding site: carbohydrate (Asn) (covalent) #sta

F:371,376/Binding site: keratan sulfate (Thr) (covalent) #status predicted

Query Match 14.2% Score 158.5; DB 1; Length 2415;
Best Local Similarity 29.1%; Pred.No.2e-05;
Matches 52; Conservative 22; Mismatches 62; Indels 43; Gaps 11;

Qy 9 QPVCRGG---TORPCYKVIVFHDTSRRLNFEEAKEACRRDGQLVLSIESDEQKLIEKFI 65
 ||| :||: |:: |:|: |:: |:|: |:: |:|: |:: |:|: |:: |:|: |:
Db 2202 QEVCCEGMNKYGHCYA--HFDP---RETWVAERRCREQQSHSLSSVTPEEQ----EFV 2252

Qy 66 ENLLPSDGDFF-WGLRRLREKQSNSTACDLAYWTGSISQFRNWYYVDSPS---CGSEVC 121
 |||: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:
Db 2253 NN-----NAQDVQMIGL-----NDRTIEDGRFWSDHPMQFENWRPNQDNPFPAAGEDC 2301

Qy 122 VVM-YHOPAPAGTGGPYMFQWNDRCNMKNFNFKYS----DEKPVPVSREAAGEETE 175
 |||: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:
Db 2302 VVMIWHEKG-----EMWDVPCNYHLPTFCCKGTVACGEPPVVEHARTFGKKD 2349

RESULT 14

B42755

E-selectin precursor - mouse

N:Alternate names: endothelial leukocyte adhesion molecule 1 (ELAM-1)

C:Species: Mus musculus (house mouse)

C>Date: 04-Mar-1993 #sequence revision 31-Dec-2000 #text_change 02-Aug-2002

C:Accession: S23174; B42755

R:Becker-Andre, M.; van Hujsdijnen, R.H.; Losberger, C.; Whelan, J.; Delamarter, J.F.
Eur. J. Biochem. 206, 401-411, 1992

A>Title: Murine endothelial leukocyte-adhesion molecule 1 is a close structural and fun

A:Reference number: S23174; MUID:92283265; PMID:1375914

A:Accession: S23174

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-612 <BEC>

A:Cross-references: GB:M80778; NID:g193014; PIDN:AAA37547.1; PID:g193015

R:Weller, A.; Isenmann, S.; Vestweber, D.

J. Biol. Chem. 267, 15176-15183, 1992

A>Title: Cloning of the mouse endothelial selectins. Expression of both E- and P-select

A:Reference number: A42755; MUID:92340571; PMID:1378846

A:Accession: B42755

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 'MKATAGV', 1-389, 391-612 <WE>

A:Cross-references: GB:M87862; NID:g193107

A:Experimental source: endothelial cells

A>Note: sequence extracted from NCBI backbone (NCBIP:109470)

A>Note: it is uncertain in GenBank entry MUSESELEC, release 117.0, (PIDN:AAA37577.1; PID:g

A:Note: it is uncertain whether the initiator is Met-1 or the AUG codon preceding that

C:Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology;

C:Keywords: glycoprotein; transmembrane protein

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:07:10 ; Search time 19.6848 Seconds
(without alignments)
1954.412 Million cell updates/sec

Title: US-09-887-855-2_COPY_22_227

Perfect score: 1115
Sequence: 1 ATGRLLSGQPCVCGTQRPC.....BEDAKTKFESREAAALNLAY 206

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 1115 | 100.0 | 206 | 9 | US-09-887-855-5 |
| 2 | 1115 | 100.0 | 374 | 9 | US-09-887-855-2 |
| 3 | 1115 | 100.0 | 374 | 12 | US-10-094-749-2090 |
| 4 | 1115 | 100.0 | 374 | 15 | US-10-149-819-15 |
| 5 | 1111 | 99.6 | 374 | 12 | US-10-094-749-2142 |
| 6 | 1101 | 98.7 | 382 | 10 | US-09-909-320-137 |
| 7 | 1101 | 98.7 | 382 | 10 | US-09-909-088B-137 |
| 8 | 1101 | 98.7 | 382 | 10 | US-09-905-291A-137 |
| 9 | 1101 | 98.7 | 382 | 10 | US-09-902-853-137 |
| 10 | 1101 | 98.7 | 382 | 10 | US-09-907-824-137 |
| 11 | 1101 | 98.7 | 382 | 10 | US-09-907-841-137 |
| 12 | 1101 | 98.7 | 382 | 11 | US-09-904-011-137 |
| 13 | 1101 | 98.7 | 382 | 11 | US-09-906-742-137 |
| 14 | 1101 | 98.7 | 382 | 11 | US-09-906-838-137 |
| 15 | 1101 | 98.7 | 382 | 11 | US-09-907-613-137 |

| | | | | | | |
|----|------|------|-----|----|--------------------|-------------------|
| 16 | 1101 | 98.7 | 382 | 11 | US-09-907-942-137 | Sequence 137, App |
| 17 | 1101 | 98.7 | 382 | 11 | US-09-904-859-137 | Sequence 137, App |
| 18 | 1101 | 98.7 | 382 | 11 | US-09-909-204-137 | Sequence 137, App |
| 19 | 1101 | 98.7 | 382 | 11 | US-09-904-820-137 | Sequence 137, App |
| 20 | 1101 | 98.7 | 382 | 11 | US-09-904-786-137 | Sequence 137, App |
| 21 | 1101 | 98.7 | 382 | 11 | US-09-906-646-137 | Sequence 137, App |
| 22 | 1101 | 98.7 | 382 | 11 | US-09-906-700-137 | Sequence 137, App |
| 23 | 1101 | 98.7 | 382 | 11 | US-09-903-786-137 | Sequence 137, App |
| 24 | 1101 | 98.7 | 382 | 11 | US-09-902-903-137 | Sequence 137, App |
| 25 | 1101 | 98.7 | 382 | 11 | US-09-903-749A-137 | Sequence 137, App |
| 26 | 1101 | 98.7 | 382 | 11 | US-09-904-119-137 | Sequence 137, App |
| 27 | 1101 | 98.7 | 382 | 11 | US-09-904-956-137 | Sequence 137, App |
| 28 | 1101 | 98.7 | 382 | 11 | US-09-902-736-137 | Sequence 137, App |
| 29 | 1101 | 98.7 | 382 | 11 | US-09-907-794-137 | Sequence 137, App |
| 30 | 1101 | 98.7 | 382 | 11 | US-09-903-943-137 | Sequence 137, App |
| 31 | 1101 | 98.7 | 382 | 11 | US-09-904-462-137 | Sequence 137, App |
| 32 | 1101 | 98.7 | 382 | 11 | US-09-907-925-137 | Sequence 137, App |
| 33 | 1101 | 98.7 | 382 | 11 | US-09-902-692-137 | Sequence 137, App |
| 34 | 1101 | 98.7 | 382 | 11 | US-09-903-520-137 | Sequence 137, App |
| 35 | 1101 | 98.7 | 382 | 11 | US-09-905-056-137 | Sequence 137, App |
| 36 | 1101 | 98.7 | 382 | 11 | US-09-909-064-137 | Sequence 137, App |
| 37 | 1101 | 98.7 | 382 | 11 | US-09-904-553-137 | Sequence 137, App |
| 38 | 1101 | 98.7 | 382 | 11 | US-09-905-381-137 | Sequence 137, App |
| 39 | 1101 | 98.7 | 382 | 11 | US-09-905-088-137 | Sequence 137, App |
| 40 | 1101 | 98.7 | 382 | 11 | US-09-907-575-137 | Sequence 137, App |
| 41 | 1101 | 98.7 | 382 | 11 | US-09-905-075-137 | Sequence 137, App |
| 42 | 1101 | 98.7 | 382 | 11 | US-09-902-759-137 | Sequence 137, App |
| 43 | 1101 | 98.7 | 382 | 11 | US-09-902-634-137 | Sequence 137, App |
| 44 | 1101 | 98.7 | 382 | 11 | US-09-902-713-137 | Sequence 137, App |
| 45 | 1101 | 98.7 | 382 | 11 | US-09-907-979-137 | Sequence 137, App |

ALIGNMENTS

RESULT 1
US-09-887-855-5
; Sequence 5, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-887-855-5

Query Match 100.0%; Score 1115; DB 9; Length 206;
Best Local Similarity 100.0%; Pred. No. 7e-105;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGRLLSGQPCVCGTQRPCYKVIYFHDTSRLNLPPEAKEACRRDGGQLVSISEDEQKL 60
Db 1 ATGRLLSGQPCVCGTQRPCYKVIYFHDTSRLNLPPEAKEACRRDGGQLVSISEDEQKL 60
QY 61 IEKFTIENLLPSDGFWIGLRRRREKQSNSTACODLYAWTDGSIQSFRNMYVDPSGSEV 120
Db 61 IEKFTIENLLPSDGFWIGLRRRREKQSNSTACODLYAWTDGSIQSFRNMYVDPSGSEV 120
QY 121 CVMVTHQPSAPAGIGGPTMFWQNDRCNNKNNFICKYSDKPAVPSPREAGEGETELTPV 180
Db 121 CVMVTHQPSAPAGIGGPTMFWQNDRCNNKNNFICKYSDKPAVPSPREAGEGETELTPV 180
QY 181 LPEETQEDAKTKFESREAAALNLAY 206
Db 181 LPEETQEDAKTKFESREAAALNLAY 206

RESULT 2
US-09-887-855-2
; Sequence 2, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-887-855-2

Query Match 100.0%; Score 1115; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.7e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRCGGTQRCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSISEDEQKL 60
DB 22 ATGRLLSGQPVCRCGGTQRCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSISEDEQKL 81

QY 61 IEKFIENLLPSDGFHWIGLRREKQSNSTACODLYAWTDGSIQSRNNWYVDEPSCGSEV 120
DB 82 IEKFIENLLPSDGFHWIGLRREKQSNSTACODLYAWTDGSIQSRNNWYVDEPSCGSEV 141

QY 121 CVVMYHQPAPAGIGPYFQWDDRCNMKNPFICKYSDEKPAVPSREAGEETEELTPV 180
DB 142 CVVMYHQPAPAGIGPYFQWDDRCNMKNPFICKYSDEKPAVPSREAGEETEELTPV 201

QY 181 LPEETOEDAKKTFKESREAAALNLAY 206
DB 202 LPEETOEDAKKTFKESREAAALNLAY 227

RESULT 3
US-10-094-749-2090
; Sequence 2090, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14

; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2090
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2090

Query Match 100.0%; Score 1115; DB 12; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.7e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRCGGTQRCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSISEDEQKL 60
DB 22 ATGRLLSGQPVCRCGGTQRCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSISEDEQKL 81

QY 61 IEKFIENLLPSDGFHWIGLRREKQSNSTACODLYAWTDGSIQSRNNWYVDEPSCGSEV 120
DB 82 IEKFIENLLPSDGFHWIGLRREKQSNSTACODLYAWTDGSIQSRNNWYVDEPSCGSEV 141

QY 121 CVVMYHQPAPAGIGPYFQWDDRCNMKNPFICKYSDEKPAVPSREAGEETEELTPV 180
DB 142 CVVMYHQPAPAGIGPYFQWDDRCNMKNPFICKYSDEKPAVPSREAGEETEELTPV 201

QY 181 LPEETOEDAKKTFKESREAAALNLAY 206
DB 202 LPEETOEDAKKTFKESREAAALNLAY 227

RESULT 4
US-10-149-819-15
; Sequence 15, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: SHAH, Purvi
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BUREFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PF-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 3143411CDI
US-10-149-819-15

Query Match 100.0%; Score 1115; DB 15; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.7e-104;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRCGGTQRCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSISEDEQKL 60
DB 22 ATGRLLSGQPVCRCGGTQRCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSISEDEQKL 81

QY 61 IEKFIENLLPSDGFHWIGLRREKQSNSTACODLYAWTDGSIQSRNNWYVDEPSCGSEV 120
DB 82 IEKFIENLLPSDGFHWIGLRREKQSNSTACODLYAWTDGSIQSRNNWYVDEPSCGSEV 141

Qy 121 CVMYHQPAPAGIGGYPYFQWDDRCNMKNFKICKYSDEKPAVPSREAGEETELTPV 180
Db 142 CVMYHQPAPAGIGGYPYFQWDDRCNMKNFKICKYSDEKPAVPSREAGEETELTPV 201
Qy 181 LPEETOEDAKTKFESREAAINLAY 206
Db 202 LPEETOEDAKTKFESREAAINLAY 227

RESULT 5

US-10-094-749-2142

; Sequence 2142, Application US/10094749

; Publication No. US20030219741A1

; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: HIO, YURI

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHICO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOKUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA

; FILE REFERENCE: 084335/0160

; CURRENT APPLICATION NUMBER: US/10/094,749

; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 60/350,435

; PRIOR FILING DATE: 2002-01-24

; PRIOR APPLICATION NUMBER: JP 2001-328381

; PRIOR FILING DATE: 2001-09-14

; NUMBER OF SEQ ID NOS: 3381

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2142

; LENGTH: 374

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-094-749-2142

Query Match

Best Local Similarity 99.6%; Score 1111; DB 12; Length 374;

Matches 205; Conservativity 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGRLSGQVPCGGTGPCVYFHDTSRRLNFEAKACRRDGGQLVSIKEDQKL 60
Db 22 ATGRLSGQVPCGGTGPCVYFHDTSRRLNFEAKACRRDGGQLVSIKEDQKL 81

Qy 61 IEKFIENLLPSDGFDFWGLRRREKQSNSTACQDLYAWTDGSIQFPRNMYVDPSGCGSEV 120
Db 82 IEKFIENLLPSDGFDFWGLRRREKQSNSTACQDLYAWTDGSIQFPRNMYVDPSGCGSEV 141

Qy 121 CVMYHQPAPAGIGGYPYFQWDDRCNMKNFKICKYSDEKPAVPSREAGEETELTPV 180
Db 142 CVMYHQPAPAGIGGYPYFQWDDRCNMKNFKICKYSDEKPAVPSREAGEETELTPV 201

Qy 181 LPEETOEDAKTKFESREAAINLAY 206
Db 202 LPEETOEDAKTKFESREAAINLAY 227

RESULT 6

US-09-909-320-137

; Sequence 137, Application US/09909320

; Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-320-137

Query Match 98.7%; Score 1101; DB 10; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQPVCRGQTQPCYKVIYFHDTSRLNPFEEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLRLRGQGPVCRGQTQPCYKVIYFHDTSRLNPFEEAKEACRRDGGQLVSI 81

QY 53 ESEDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVD 112
Db 82 ESEDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVD 141

QY 113 EPSCGSEVCVMYHQPAPAGIGGPMYFQWDDRCNMKNFICKYDDEKPAVPSRAEAGE 172
Db 142 EPSCGSEVCVMYHQPAPAGIGGPMYFQWDDRCNMKNFICKYDDEKPAVPSRAEAGE 201

QY 173 ETELTTPVLPEETQEDAKKTFKESREAAALNLAY 206
Db 202 ETELTTPVLPEETQEDAKKTFKESREAAALNLAY 235

RESULT 7

US-09-909-088B-137
; Sequence 137, Application US/09909088B
; Patent No. US20020146709A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertitsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,088B
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089

Query Match 98.7%; Score 1101; DB 10; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQPVCRGQTQPCYKVIYFHDTSRLNPFEEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLRLRGQGPVCRGQTQPCYKVIYFHDTSRLNPFEEAKEACRRDGGQLVSI 81

QY 53 ESEDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVD 112
Db 82 ESEDEQKLEKFIENLLPSDGFWIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVD 141

QY 113 EPSCGSEVCVMYHQPAPAGIGGPMYFQWDDRCNMKNFICKYDDEKPAVPSRAEAGE 172
Db 142 EPSCGSEVCVMYHQPAPAGIGGPMYFQWDDRCNMKNFICKYDDEKPAVPSRAEAGE 201

QY 173 ETELTTPVLPEETQEDAKKTFKESREAAALNLAY 206
Db 202 ETELTTPVLPEETQEDAKKTFKESREAAALNLAY 235

RESULT 8

US-09-905-291A-137
; Sequence 137, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertitsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel

```

; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-905-291A-137

Query Match          98.7%; Score 1101; DB 10; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSI 52
Db 22 ATGRLLSASDLRLGGQVCRGGTQPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSI 81
Qy 53 ESEDEQKLEKFIENLLPSDGFVIGLRRRREKQSNSTACQDLYAWTDGSGISQFRNWTVD 112
Db 82 ESEDEQKLEKFIENLLPSDGFVIGLRRRREKQSNSTACQDLYAWTDGSGISQFRNWTVD 141
Qy 113 EPSCGSEVCVMYHQPSAPAGIGPYFMQWDDRCNKNQNFICKYSDEKPAVPSREAGE 172
Db 142 EPSCGSEVCVMYHQPSAPAGIGPYFMQWDDRCNKNQNFICKYSDEKPAVPSREAGE 201
Qy 173 ETELTTPVLPEETOEDAKKTFKESREAAALNLAY 206
Db 202 ETELTTPVLPEETOEDAKKTFKESREAAALNLAY 235

RESULT 9
US-09-902-853-137
; Sequence 137, Application US/09902853
; Publication No. US20020192659A1

; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/902,853
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US/09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-902-853-137

Query Match          98.7%; Score 1101; DB 10; Length 382;
```



```

; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,841
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-907-841-137
```

```

Query Match      98.7%; Score 1101; DB 10; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSI 52
DB 22 ATGRLLSASDLDRGGQVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSI 81

QY 53 ESEDEQKLIKFIENLLPSDGFWDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVD 112
DB 82 ESEDEQKLIKFIENLLPSDGFWDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVD 141

QY 113 EPCGSEVCMVYHQPAPAGIGGYPYFQWDDRCNKNKFNICKYDEKPAVPSREAGE 172
DB 142 EPCGSEVCMVYHQPAPAGIGGYPYFQWDDRCNKNKFNICKYDEKPAVPSREAGE 201

QY 173 ETELTTPVLPETQEDAKTKFKESRAALNLAY 206
DB 202 ETELTTPVLPETQEDAKTKFKESRAALNLAY 235
```

```

RESULT 12
US-09-904-011-137
; Sequence 137, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
```

```

; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
; ORGANISM: Homo Sapien
US-09-904-011-137
```

```

Query Match      98.7%; Score 1101; DB 11; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSI 52
DB 22 ATGRLLSASDLDRGGQVCRGGTQPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSI 81

QY 53 ESEDEQKLIKFIENLLPSDGFWDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVD 112
DB 82 ESEDEQKLIKFIENLLPSDGFWDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVD 141
```

QY 113 EPSCGSEVCVMYHQSAPAGIGGPFYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGE 172
Db 142 EPSCGSEVCVMYHQSAPAGIGGPFYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGE 201
QY 173 ETELTPVLPEETQEEDAKKTFKESREAAALNLAY 206
Db 202 ETELTPVLPEETQEEDAKKTFKESREAAALNLAY 235

RESULT 13

US-09-906-742-137
; Sequence 137, Application US/09906742
; Publication No. US20030023054A1

; GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; TITLE OF INVENTION: Acids Encoding the Same

; FILE REFERENCE: 10466-14

; CURRENT APPLICATION NUMBER: US/09/906,742

; PRIOR FILING DATE: 2001-07-16

; PRIOR APPLICATION NUMBER: 09/665,350

; PRIOR FILING DATE: 2000-09-18

; PRIOR APPLICATION NUMBER: PCT/US00/04414

; PRIOR FILING DATE: 2000-02-22

; PRIOR APPLICATION NUMBER: US 60/143,048

; PRIOR FILING DATE: 1999-07-07

; PRIOR APPLICATION NUMBER: US 60/145,698

; PRIOR FILING DATE: 1999-07-26

; PRIOR APPLICATION NUMBER: US 60/146,222

; PRIOR FILING DATE: 1999-07-28

; PRIOR APPLICATION NUMBER: PCT/US99/20594

; PRIOR FILING DATE: 1999-09-08

; PRIOR APPLICATION NUMBER: PCT/US99/20944

; PRIOR FILING DATE: 1999-09-13

; PRIOR APPLICATION NUMBER: PCT/US99/21090

; PRIOR FILING DATE: 1999-09-15

; PRIOR APPLICATION NUMBER: PCT/US99/21547

; PRIOR FILING DATE: 1999-09-15

; PRIOR APPLICATION NUMBER: PCT/US99/23089

; PRIOR FILING DATE: 1999-10-05

; PRIOR APPLICATION NUMBER: PCT/US99/28214

; PRIOR FILING DATE: 1999-11-29

; PRIOR APPLICATION NUMBER: PCT/US99/28313

; PRIOR FILING DATE: 1999-11-30

; PRIOR APPLICATION NUMBER: PCT/US99/28564

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/28565

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/28565

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/28565

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/28565

; PRIOR FILING DATE: 1999-12-02

; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-906-742-137

Query Match

Best Local Similarity 98.7%; Score 1101; DB 11; Length 382;

Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLIS-----CQVCRGCTGTPCYKVYFPHDTSRLNPFEEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLIRGGQPVCRGCTGTPCYKVYFPHDTSRLNPFEEAKEACRRDGGQLVSI 81
QY 53 ESEDEQKLIETKFIENLLPSDGFWIGLRRREKQSNSTACQDLYANTDGSISQFRNMYVD 112
Db 82 ESEDEQKLIETKFIENLLPSDGFWIGLRRREKQSNSTACQDLYANTDGSISQFRNMYVD 141
QY 113 EPSCGSEVCVMYHQSAPAGIGGPFYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGE 172
Db 142 EPSCGSEVCVMYHQSAPAGIGGPFYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGE 201
QY 173 ETELTPVLPEETOEDAKKTFKESREAAALNLAY 206
Db 202 ETELTPVLPEETOEDAKKTFKESREAAALNLAY 235

RESULT 14

US-09-906-838-137
; Sequence 137, Application US/09906838
; Publication No. US20030027143A1

; GENERAL INFORMATION:

; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/906,838
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18

;; PRIOR APPLICATION NUMBER: PCT/US00/04414
;; PRIOR FILING DATE: 2000-02-22
;; PRIOR APPLICATION NUMBER: US 60/143,048
;; PRIOR FILING DATE: 1999-07-07
;; PRIOR APPLICATION NUMBER: US 60/145,698
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: US 60/146,222
;; PRIOR FILING DATE: 1999-07-28
;; PRIOR APPLICATION NUMBER: PCT/US99/20594
;; PRIOR FILING DATE: 1999-09-08
;; PRIOR APPLICATION NUMBER: PCT/US99/20944
;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/21547
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/23089
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: PCT/US99/28214
;; PRIOR FILING DATE: 1999-11-29
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: 1999-11-30
;; PRIOR APPLICATION NUMBER: PCT/US99/28564
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/28565
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: PCT/US99/30911
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US99/30999
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US00/00219
;; PRIOR FILING DATE: 2000-01-05
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 137
;; LENGTH: 382
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-906-838-137

Query Match 98.7%; Score 1101; DB 11; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSI 81

Qy 53 ESEDEQKLIKFIENLLPSDGFNIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVD 112
Db 82 ESEDEQKLIKFIENLLPSDGFNIGLRRREKQSNSTACODLYAWTDGSIQFRNMYVD 141

Qy 113 EPCSGSEVCVMYHOPAPAGIGPYFQNDRCNNKNNFICKYSEKPAVPSREAGE 172
Db 142 EPCSGSEVCVMYHOPAPAGIGPYFQNDRCNNKNNFICKYSEKPAVPSREAGE 201

Qy 173 ETELTPVLPEETOEDAKTKFESREAAALNLAY 206
Db 202 ETELTPVLPEETOEDAKTKFESREAAALNLAY 235

RESULT 15
US-09-907-613-137
; Sequence 137, Application US/09907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone

;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Fong, Sherman
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerber, Hanspeter
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, A.
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Grimaldi, Christopher J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Hillan, Kenneth, J.
;; APPLICANT: Kljavin, Ivar J.
;; APPLICANT: Mather, Jennie P.
;; APPLICANT: Pan, James
;; APPLICANT: Paoni, Nicholas F.
;; APPLICANT: Roy, Margaret Ann
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Williams, P. Mickey
;; APPLICANT: Wood, William, I.
;; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;; FILE REFERENCE: 10466-14
;; CURRENT APPLICATION NUMBER: US/09/907,613
;; PRIOR FILING DATE: 2001-07-17
;; PRIOR APPLICATION NUMBER: PCT/US00/04414
;; PRIOR FILING DATE: 2000-02-22
;; PRIOR APPLICATION NUMBER: US 60/143,048
;; PRIOR FILING DATE: 1999-07-07
;; PRIOR APPLICATION NUMBER: US 60/145,698
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: US 60/146,222
;; PRIOR FILING DATE: 1999-07-28
;; PRIOR APPLICATION NUMBER: PCT/US99/20594
;; PRIOR FILING DATE: 1999-09-08
;; PRIOR APPLICATION NUMBER: PCT/US99/20944
;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/21547
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/23089
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: PCT/US99/28214
;; PRIOR FILING DATE: 1999-11-29
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: 1999-11-30
;; PRIOR APPLICATION NUMBER: PCT/US99/28564
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/28565
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: PCT/US99/30911
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US99/30999
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US00/00219
;; PRIOR FILING DATE: 2000-01-05
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 137
;; LENGTH: 382
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-907-613-137

Query Match 98.7%; Score 1101; DB 11; Length 382;
Best Local Similarity 96.3%; Pred. No. 4.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSI 81

| | | | | | | | | | | | | | | |
|----|-----|-----------|-------|-------|-------|-------|------|-------|------|------|------|------|------|-----|
| Qy | 53 | ESEDEQKLI | EKFIE | NLLPS | DGDFW | IGLR | RREK | QSNST | ACOD | LYAW | TGSI | QFRN | MYVD | 112 |
| Db | 82 | ESEDEQKLI | EKFIE | NLLPS | DGDFW | IGLR | RREK | QSNST | ACOD | LYAW | TGSI | QFRN | MYVD | 141 |
| Qy | 113 | EPSCGSE | VVVM | YHQP | SAPAG | IGGYP | MFQW | NDRC | NMKN | NFKY | SD | CKPA | VP | 172 |
| Db | 142 | EPSCGSE | VVVM | YHQP | SAPAG | IGGYP | MFQW | NDRC | NMKN | NFKY | SD | CKPA | VP | 201 |
| Qy | 173 | ETELT | TPVL | PEETO | EDAK | TFKS | REAA | NLAY | 206 | | | | | |
| Db | 202 | ETELT | TPVL | PEETO | EDAK | TFKS | REAA | NLAY | 235 | | | | | |

Search completed: December 22, 2003, 16:16:47
Job time : 20.6848 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:05:09 ; Search time 11.9128 Seconds
(without alignment)
1476.391 Million cell updates/sec

Title: US-09-887-855-2

Perfect score: 2000
Sequence: 1 MRPGTALQAVLLAVLLVGLR.....PDQMGRSKESQWVENEIYGY 374

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 657 | 32.9 | 273 | 1 | Q9cxm0 mus musculus |
| 2 | 645 | 32.2 | 273 | 1 | CHOD_MOUSE |
| 3 | 183 | 9.2 | 1456 | 1 | MANR_HUMAN |
| 4 | 177 | 8.8 | 1268 | 1 | PGCN_MOUSE |
| 5 | 176 | 8.8 | 1321 | 1 | PGCN_HUMAN |
| 6 | 174.5 | 8.7 | 3381 | 1 | PGCV_BOVIN |
| 7 | 174 | 8.7 | 1257 | 1 | PGCN_RAT |
| 8 | 174 | 8.7 | 2738 | 1 | PGCV_RAT |
| 9 | 174 | 8.7 | 338 | 1 | PGCV_MOUSE |
| 10 | 172.5 | 8.6 | 643 | 1 | PGCV_HUMAN |
| 11 | 172.5 | 8.6 | 643 | 1 | CD93_RAT |
| 12 | 171 | 8.6 | 3562 | 1 | PGCV_CHICK |
| 13 | 158.5 | 7.9 | 2364 | 1 | PGCA_BOVIN |
| 14 | 158.5 | 7.9 | 2415 | 1 | PGCA_HUMAN |
| 15 | 157.5 | 7.9 | 652 | 1 | CD93_HUMAN |
| 16 | 155.5 | 7.8 | 2333 | 1 | PGCA_CANFA |
| 17 | 154.5 | 7.7 | 612 | 1 | LEM2_MOUSE |
| 18 | 154.5 | 7.7 | 644 | 1 | LECG_MOUSE |
| 19 | 153.5 | 7.7 | 158 | 1 | LECG_TRIST |
| 20 | 153.5 | 7.7 | 2124 | 1 | PGCA_RAT |
| 21 | 153 | 7.6 | 162 | 1 | LECG_MEGRO |
| 22 | 152 | 7.6 | 912 | 1 | PGCB_BOVIN |
| 23 | 151 | 7.5 | 197 | 1 | CLFI_HUMAN |
| 24 | 149 | 7.4 | 321 | 1 | FCE2_HUMAN |
| 25 | 148.5 | 7.4 | 152 | 1 | IXA_TRIFL |
| 26 | 148.5 | 7.4 | 2132 | 1 | PGCA_MOUSE |
| 27 | 147 | 7.3 | 893 | 1 | PGCB_MOUSE |
| 28 | 146 | 7.2 | 2109 | 1 | PGCA_CHICK |
| 29 | 145 | 7.2 | 893 | 1 | PGCB_RAT |
| 30 | 144.5 | 7.2 | 173 | 1 | LECG_MEGRO |
| 31 | 144.5 | 7.2 | 372 | 1 | LEM1_RAT |
| 32 | 143.5 | 7.2 | 372 | 1 | LEM1_MOUSE |
| 33 | 143 | 7.1 | 549 | 1 | LEM2_RAT |
| 34 | 141.5 | 7.1 | 165 | 1 | LIT1_MOUSE |
| 35 | 141.5 | 7.1 | 331 | 1 | FCE2_MOUSE |
| 36 | 140 | 7.0 | 370 | 1 | LEM1_BOVIN |
| 37 | 139 | 7.0 | 248 | 1 | PSPA_HUMAN |
| 38 | 139 | 7.0 | 283 | 1 | LECA_SARPE |
| 39 | 138.5 | 6.9 | 172 | 1 | LECA_PLEWA |
| 40 | 138.5 | 6.9 | 202 | 1 | TETN_MOUSE |
| 41 | 137.5 | 6.9 | 372 | 1 | LEM1_MACMU |
| 42 | 137.5 | 6.9 | 372 | 1 | LEM1_PAPHA |
| 43 | 136 | 6.8 | 175 | 1 | LITH_BOVIN |
| 44 | 134.5 | 6.7 | 132 | 1 | ACAL_ANGAN |
| 45 | 134.5 | 6.7 | 372 | 1 | LEM1_PONPY |

ALIGNMENTS

RESULT 1

CHOD_MOUSE

ID CHOD_MOUSE STANDARD; PRT; 273 AA.

AC Q9CXMO: Q8VI31; 127863

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Chondrolectin precursor (Transmembrane protein MT75).

GN CHODL

OS Mus musculus (Mouse)

OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RA Weng L., Smits P., Hubner R., Wouters J., Merregaert J.;

RT "Mt75, a low expressed c-type lectin gene involving in

chondrogenesis."

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

RR [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Embryonic head;

MEMLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

Kadoca K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

Fleischmann W., Gaasterland T., Giesi C., King B., Kochiwa H.,

Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,

Schriml L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,

Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

Brustein M.J., Bult C., Bojunga N., Carninci P., de Bona M.F.,

Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,

Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,

Hayashizaki Y.;

RL "Functional annotation of a full-length mouse cDNA collection.;"

RR Nature 409:685-690 (2001).

CC - SUBCELLULAR LOCATION: Type I membrane protein (Potential).

CC - SIMILARITY: Contains 1 C-type lectin family domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).

DR EMBL; AF311699; AAL50354.1; -


```

DR EMBL; AK012455; BAB29226.1; -.
DR HSSP; P22897; IEGG.
DR MGD; MGI:2179069; Chodl.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C TYPE LECTIN 1; FALSE_NEG.
DR PROSITE; PS00041; C TYPE LECTIN 2; 1. Signal.
KW Lectin; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 273 CHONDROLECTIN.
FT DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 217 237 POTENTIAL.
FT DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 179 C-TYPE LECTIN.
FT CARBOHYD 86 86 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 24 24 V -> W (IN REF. 2).
FT CONFLICT 179 179 T -> K (IN REF. 2).
SQ SEQUENCE 273 AA; 30303 MW; E052D933F244F4C7 CRC64;

Query Match 32.98; Score 657; DB 1; Length 273;
Best Local Similarity 49.18; Pred. No. 3.4e-48;
Matches 139; Conservative 38; Mismatches 80; Indels 26; Gaps 9;

QY 10 VLLAVLLVGLRAATGRLLSQGVPCRGCTGTPCYKVIYFHDTSRLNPEEAKACRRDGGQ 69
DB 8 LLGALLCAOGAFARRVVSQKVCFAADVHPCYKMAVFHLSRVSFQEARLACESEGV 67
QY 70 LVISIEDEOKLIEKFIENLLP-----SDGDFWIGLARREKQNSNACQDLVATWDGSI 124
DB 68 LLSLENAEQKLIESMLQNLTKPTGTISDGFWIGLARSGDGT-SGACPDLYQWSDGSS 126
QY 125 SQRNMYVDPSGCGSEVGVVYHOPSPAGIGGPGYMFQWNRDRCNMKNKNIKYSDE-KP 183
DB 127 SQRNMYVDPSGCGSEKCVVYHQPNTANGLGGPYLQWNRDRCNMKNKNIYEPYEPHP 186
QY 184 AVPSREAEGRETELTPVLPEETQEDAKTKFKESREAAAL--NLAYLILPSILLLLIV 241
DB 187 TEPA-----EKPYLTNQ--PEETHENVV-----VTEAGIIPNLIVITPILLLLIV 233
QY 242 TTVVGVWVICRKKK-REQPPSTKQHTIWPSPHQGNSPDLEV 283
DB 234 ALGTCCCFQMLHKGSRKTSFN---QSTLWISKSTRKSGMEV 273

RESULT 2
CHOD_HUMAN STANDARD; PRT; 273 AA.
AC Q9H9P2; Q9HCY3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Chondrolectin precursor (Transmembrane protein MT75) (PRED12 protein).
GN CHODL OR C21ORF68.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=2074930; PubMed=12079284;
RA Weng L., Smits P., Wauters J., Merregaert J.;
RT "Molecular cloning and characterization of human chondrolectin, a novel type I transmembrane protein homologous to C-type lectins."
RL Genomics 80:62-70(2002).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA Soeda E., Onki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,

```

```

RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuayama S., Antonarakis S.E.,
RA Minoashima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Schaefer M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21."
RL Nature 405:311-319(2000).
[3]
RN SEQUENCE FROM N.A.
RX TISSUE=Brain;
MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haiech F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udutin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[4]
RN SEQUENCE OF 27-273 FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Wagatsuma M., Hosoi R., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN TISSUE SPECIFICITY.
RX MEDLINE=21564202; PubMed=11707072;
RA Raymond A., Friedli M., Neergard Henriksen C., Chapot F., Deutsch S.,
RA Ucla C., Roessier C., Lyle R., Guipponi M., Antonarakis S.E.;
RT "From PREDS and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map."
RL Genomics 78:46-54(2001).
CC -I- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -I- TISSUE SPECIFICITY: Found in spleen, testis, prostate and fetal liver. Expression limited to vascular muscle of testis, smooth muscle of prostate stroma, heart muscle, skeletal muscle, crypts of small intestine, and red pulp of spleen.
CC -I- PTM: N-glycosylated.
CC -I- SIMILARITY: Contains 1 C-type lectin family domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).
-----
CC EMBL; AF257472; AAL05981.1; -.
CC EMBL; AL163217; CAB90388.1; -.
CC EMBL; BC009418; AAH09418.1; -.
CC EMBL; AK022689; BAB14181.1; ALT_INIT.

```

```

DR HSP: P22897; 1EGG.
DR Genew; HGNC:17807; CHODL.
DR MIM; 607247; -.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CUECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; FALSE_NEG.
DR PROSITE; PS00441; C-TYPE LECTIN 2; 1.
KW Lectin; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 273 CHONDROLECTIN.
FT DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 217 237 POTENTIAL.
FT DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 179 C-TYPE LECTIN.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 273 AA; 30431 MW; F490A0AFB572A311 CRC64;

Query Match 32.2%; Score 645; DB 1; Length 273;
Best Local Similarity 48.4%; Pred. No. 3.6e-47;
Matches 137; Conservative 40; Mismatches 80; Indels 26; Gaps 9;

QY 10 VLLAVLLGLRAATGRLLSGQPCVCGGTQPCYKVIYFHTSRLNFEAKACRRDGGQ 69
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8 LGAALLCGHGFACRRVSGQKCFADFKHPCYKWAYFHELSRVSFQEARLACESEGV 67

QY 70 LVSISEDEQKLIKFIEINLLP-----SDGFWIGLRREREKQSNSTACODLYAWTDGSI 124
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 LLSLENEAEOKLIESMLQNLTPGTVISDGDGFWIGLWRNGDQT-SGACPDLYQWSDGNS 126

QY 125 SQFRNWWYDEPSCGSEVCVVMYHQSPAGIGPYFMQWDDRCNMKNFKICKYSDE-KP 183
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
127 SQYRNWYTDPSGSEKCVVMYHQTPANPLGPGVLYQWDDRCNMKNYICKYPEINP 186

QY 184 AVPSREAGEETELTPVLPEETQEDAKTFKESREAL--NLAYILPSIPLLLLVV 241
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187 TAPV-----EKPLTNQ--PGDTHQNV-----VTEAGITPNLYVVIPTLELLLV 233

QY 242 TTVVCVWVICRRK-REQDPSTKQHTIWPSPHOGNSPDLEV 283
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 AFGTCFCFQMLHKSGRKTKTSPN---QSLTWLSKSTRKESGNEV 273

RESULT 3
MANR_HUMAN
ID MANR_HUMAN STANDARD; PRT; 1456 AA.
AC P22897;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Macrophage mannose receptor precursor (MMR) (CD206 antigen).
GN MRC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP TISSUE=Placenta;
RX MEDLINE=90324192; PubMed=2373685;
RA Taylor M.E., Conary J.T., Lennartz M.R., Stahl P.D., Drickamer K.;
RT "Primary structure of the mannose receptor contains multiple motifs
resembling carbohydrate-recognition domains.";
RL J. Biol. Chem. 265:12156-12162(1990).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=93052405; PubMed=1294118;
RX Km S.J., Ruiz N., Bezouska K., Drickamer K.;
RT "Organization of the gene encoding the human macrophage mannose
receptor (MRC1).";
RL Genomics 14:721-727(1992).
[3]
RN STUDIES ON THE BINDING OF INDIVIDUAL LECTIN DOMAINS.
RP

```

```

RX MEDLINE=92112893; PubMed=1730714;
RA Taylor M.E., Bezouska K., Drickamer K.;
RT "Contribution to ligand binding by multiple carbohydrate-recognition
domains in the macrophage mannose receptor.";
RL J. Biol. Chem. 267:1719-1726(1992).
[4]
RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 642-788.
RX MEDLINE=20347275; PubMed=10779515;
RA Feinberg H., Park-Snyder S., Kolatkar A.R., Heise C.T., Taylor M.E.,
Weis W.I.;
RT "Structure of a C-type carbohydrate recognition domain from the
macrophage mannose receptor.";
RL J. Biol. Chem. 275:21539-21548(2000).
CC -!- FUNCTION: MEDIATES THE ENDOCYTOSIS OF GLYCOPROTEINS BY
MACROPHAGES, IN SEVERAL RECOGNITION AND UPTAKE PROCESSES.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: CRDS 1-3 HAVE AT MOST VERY WEAK AFFINITY FOR
CARBOHYDRATE. CRD 4 SHOWS THE HIGHEST AFFINITY BINDING AND HAS
MULTISPECIFICITY FOR A VARIETY OF MONOSACCHARIDES. AT LEAST 3 CRDS
(4, 5, AND 7) ARE REQUIRED FOR HIGH AFFINITY BINDING AND
ENDOCYTOSIS OF MULTIVALENT GLYCOCONGUGATES.
CC -!- SIMILARITY: Contains 8 C-type lectin family domains.
CC -!- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -!- DATABASE: NAME=PROW; NOTE=PROW 2:85-89(2001);
WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1644341535_g.htm"
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; J05550; AAA59868.1; -
DR EMBL; M93221; AAA60389.1; -
DR EMBL; M93192; AAA60389.1; JOINED.
DR EMBL; M93193; AAA60389.1; JOINED.
DR EMBL; M93194; AAA60389.1; JOINED.
DR EMBL; M93195; AAA60389.1; JOINED.
DR EMBL; M93196; AAA60389.1; JOINED.
DR EMBL; M93197; AAA60389.1; JOINED.
DR EMBL; M93198; AAA60389.1; JOINED.
DR EMBL; M93199; AAA60389.1; JOINED.
DR EMBL; M93200; AAA60389.1; JOINED.
DR EMBL; M93201; AAA60389.1; JOINED.
DR EMBL; M93202; AAA60389.1; JOINED.
DR EMBL; M93203; AAA60389.1; JOINED.
DR EMBL; M93204; AAA60389.1; JOINED.
DR EMBL; M93205; AAA60389.1; JOINED.
DR EMBL; M93206; AAA60389.1; JOINED.
DR EMBL; M93207; AAA60389.1; JOINED.
DR EMBL; M93208; AAA60389.1; JOINED.
DR EMBL; M93209; AAA60389.1; JOINED.
DR EMBL; M93210; AAA60389.1; JOINED.
DR EMBL; M93211; AAA60389.1; JOINED.
DR EMBL; M93212; AAA60389.1; JOINED.
DR EMBL; M93213; AAA60389.1; JOINED.
DR EMBL; M93214; AAA60389.1; JOINED.
DR EMBL; M93215; AAA60389.1; JOINED.
DR EMBL; M93216; AAA60389.1; JOINED.
DR EMBL; M93217; AAA60389.1; JOINED.
DR EMBL; M93218; AAA60389.1; JOINED.
DR EMBL; M93219; AAA60389.1; JOINED.
DR EMBL; M93220; AAA60389.1; JOINED.
DR PIR; A36563; A36563.
DR PDB; 1EGG; 30-AUG-00.
DR PDB; 1EGI; 30-AUG-00.
DR Genew; HGNC:7228; MRC1.
DR MIM; 153618; -.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0005537; F:mannose binding activity; TAS.
DR GO; GO:0004872; F:receptor activity; TAS.

```

[illegible]

DR InterPro; IPR003006; Ig MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.
 DR PRINTS; PR01265; LINKMODULE.
 DR PRINTS; PR00356; ANTIFREEZEIII.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASK_HYDROXYL; 1.
 DR PROSITE; PS00615; C-TYPE_LLECTIN_1; 1.
 DR PROSITE; PS00041; C-TYPE_LLECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS00835; IG_LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
 KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 1268
 FT DOMAIN 37 157
 FT DOMAIN 158 253
 FT DOMAIN 259 355
 FT DOMAIN 960 996
 FT DOMAIN 998 1034
 FT DOMAIN 1036 1165
 FT DOMAIN 1166 1224
 FT DISULFID 58 139
 FT DISULFID 181 252
 FT DISULFID 205 226
 FT DISULFID 279 354
 FT DISULFID 303 324
 FT DISULFID 964 975
 FT DISULFID 969 984
 FT DISULFID 986 995
 FT DISULFID 1040 1051
 FT DISULFID 1068 1162
 FT DISULFID 1136 1152
 FT DISULFID 1167 1210
 FT DISULFID 1196 1223
 FT CARBOHYD 121 121
 FT CARBOHYD 339 339
 FT CARBOHYD 742 742
 FT CARBOHYD 978 978
 FT CARBOHYD 1175 1175
 FT SEQUENCE 1268 AA; 137200 MW; 3014BE202A2FAEC CRC64;
 Query Match 8.8%; Score 177; DB 1; Length 1268;
 Best Local Similarity 31.5%; Pred. No. 1e-06;
 Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;
 QY 38 ORCYKVIYFHDTSRLNFEAKACRRDGGQLVSEDEQKLEIKFIENLIPSGDFW 97
 Db 1048 QHCYR--YF---AHRRAWEADRCRRRAGHLTSVHSPEEHKFINSF-----GHNSW 1096
 QY 98 IGLRRREKOSNSTACQDLVAWTDGSGISQFRNMYVDEPS---CGSEVCVVMYHQPSAPAG 154
 Db 1097 IGLNDRTRVED-----FQWTDNTGLQYENREKQPDNFFAGGDCVVMVAHESG--- 1145
 QY 155 IGGPYMFQWMDRCNKNKPFICK 177
 Db 1146 -----RWNDVPCNLYLVVCK 1161

RESULT 5
 PGCN_HUMAN
 ID_PGCN_HUMAN STANDARD; PRT; 1321 AA.
 AC 014594; QSPK6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Neurocan core protein precursor (Chondroitin sulfate proteoglycan 3).
 GN CSPG3 OR NCAN OR NEUR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95013874; PubMed=9795216;
 RA Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Lennon G.G.;
 RT "Characterization of the human neurocan gene, CSPG3.";
 RN Gene 221:199-205(1998).
 RP [2]
 RP SEQUENCE OF 1-990 AND 1007-1321 FROM N.A.
 RA Lanerding J.E., McCready P.M., Skowronski E., Adamson A.W.,
 RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stiilwagen S.,
 RA Phan H., Velasco N., Do L., Regalia W., Terry A., Barnes J.,
 RA Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
 RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
 RA Cosfield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
 RA Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
 RA Kobayashi A., Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of an ~1 Mb region containing the MEF2B gene in
 19p12.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
 development by binding to neural cell adhesion molecules (NC-CAM
 and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
 acid.
 CC -!- TISSUE SPECIFICITY: Brain.
 CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -!- SIMILARITY: Contains 2 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 link domains.
 CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF026547; AAC80576.1; -;
 DR EMBL; AC003110; AAC86655.1; -;
 DR EMBL; AC005254; AAC25581.1; -;
 DR HSSP; P00740; LEDM
 DR Genew; HGNC:2465; CSPG3.
 DR MM; 600826; -;
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.

```

DR PRINTS; PRO1265; LINKMODULE.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCF; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF CA; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX HYDROXYL; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1321
FT DOMAIN 38 153
FT DOMAIN 159 254
FT DOMAIN 260 356
FT DOMAIN 1008 1044
FT DOMAIN 1046 1082
FT DOMAIN 1084 1213
FT DOMAIN 1214 1272
FT DISULFID 59 140
FT DISULFID 182 253
FT DISULFID 206 227
FT DISULFID 280 355
FT DISULFID 304 325
FT DISULFID 1012 1023
FT DISULFID 1017 1032
FT DISULFID 1034 1043
FT DISULFID 1088 1099
FT DISULFID 1116 1208
FT DISULFID 1184 1200
FT DISULFID 1215 1258
FT DISULFID 1244 1271
FT CARBOHYD 122 122
FT CARBOHYD 340 340
FT CARBOHYD 1026 1026
FT CARBOHYD 1223 1223
FT CONFLICT 1234 1234
FT CONFLICT 1254 1254
FT CONFLICT 1282 1282
SQ SEQUENCE 1321 AA; 142972 MW; 2EF47F823DB98088 CRC64;

Query Match 8.8%; Score 176; DB 1; Length 1321;
Best Local Similarity 31.5%; Pred. No. 1.3e-06;
Matches 45; Conservative 18; Mismatches 48; Indels 32; Gaps 6;

QY 38 QPCKYVYFHTSRRLNFEAEKACRRGGQLVLSIESDEOKLIBKFTIENLLPSGDWF 97
DB 1096 QGHCYR--YF--AHRRWEDAEDKCRRSGLTTSVHSPEHSFNSF-----GHENTW 1144

QY 98 IGLRREERQKSNSTACQDLVATWDGSIQFRWYVDERS---CGSRVCVVMYHQSPAG 154
DB 1145 IGLNDRIVERD-----FQWTDNTGLQFENWRENQPNFPAGGEDCVVMVAHESG---- 1193

QY 155 IGGPYMFWNDRCNMKNFPICK 177
DB 1194 -----RNDVPCNYPVUCK 1209

RESULT 6
PCGV_BOVIN
ID PGCV_BOVIN STANDARD; PRT; 3381 AA.
AC P81282; O77609; O77610; O77611; O77612;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)

```

```

DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP).
GN CP5G2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORMS VO; V1; V2 AND V3).
RP TISSUE=Forebrain;
RC MEDLINE=98288320; PubMed=9624174;
RX Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
KL J. Biol. Chem. 273:15758-15764(1998).
RN [2]
RN SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331
RN AND 342-348.
RP TISSUE=Spinal cord;
RC MEDLINE=92062692; PubMed=1720020;
RX Perides G., Biviano F., Bignami A.;
RA "Interaction of a brain extracellular matrix protein with hyaluronic
RA acid.";
RL Biochim. Biophys. Acta 1075:248-258(1991).
CC -1- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -1- SUBUNIT: Interacts with FBLN1 (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=P81282-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=P81282-2; Sequence=VSP_003078, VSP_003079;
CC Name=V2;
CC IsoId=P81282-3; Sequence=VSP_003080;
CC Name=V3;
CC IsoId=P81282-4; Sequence=VSP_003078, VSP_003081;
CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed
CC in the central nervous system, and in a number of mesenchymal and
CC epithelial tissues; the major isoform V2 is restricted to the
CC central nervous system.
CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (By similarity).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 2 link domains.
CC -1- SIMILARITY: Contains 2 EGF-like domains.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF060456; AAC24358.1; -
CC EMBL; AF060457; AAC24359.1; -
CC EMBL; AF060458; AAC24360.1; -
CC EMBL; AF060459; AAC24361.1; -
CC PIR; T14274; T14274.
CC PIR; T42389; T42389.
CC HSSP; P01132; 1EPG.
CC InterPro; IPR000152; Asx hydroxyl.
CC InterPro; IPR000742; EGF_2.

```


RA Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.,
RA Margolis R.U., Grumet M.;
RT "The neuronal chondroitin sulfate proteoglycan neurocan binds to the
RT neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits
RT neuronal adhesion and neurite outgrowth.";
RL J. Cell Biol. 125:669-680(1994).
CC -!- FUNCTION: May modulate neuronal adhesion and neurite growth during
CC development by binding to neural cell adhesion molecules (NG-CAM
CC and N-CAM). Chondroitin sulfate proteoglycan; binds to hyaluronic
CC acid.
CC -!- TISSUE SPECIFICITY: EARLY POSTNATAL AND ADULT BRAIN; NOT EXPRESSED
CC IN KIDNEY, LUNG, LIVER AND MUSCLE.
CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -!- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
CC DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL
CC BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M97161; AAC37679.1; -.
CC PIR: S28764; S28764.
CC HSPR: P00740; 1EDM.
DR InterPro: IPR002353; AntifreezeII.
DR InterPro: IPR000152; Asx hydroxyl.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003006; Ig MHC.
DR InterPro: IPR001304; Lectin_C.
DR InterPro: IPR000538; Linkin_C.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF00047; Ig; 1.
DR Pfam: PF00059; lectin_c; 1.
DR Pfam: PF00084; sushi_1.
DR Pfam: PF00193; Xlink; 2.
DR PRINTS: PR01265; LINKMODULE.
DR PRINTS: PR00356; ANTIFREEZEII.
DR PRODOM: PD000918; Link; 2.
DR SMART: SM00032; CCP; 1.
DR SMART: SM00034; CLECT; 1.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00409; IG; 1.
DR SMART: SM00445; LINK; 2.
DR PROSITE: PS00010; ASX HYDROXYL; 1.
DR PROSITE: PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE: PS00041; C-TYPE LECTIN_2; 1.
DR PROSITE: PS00022; EGF_1_3.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS00835; IG-LIKE; 1.
DR PROSITE: PS01241; LINK; 2.
DR Glycoprotein: Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1257 NEUROCAN CORE PROTEIN.
FT CHAIN 639 1257 150 kDa ADULT CORE GLYCOPROTEIN.
FT DOMAIN 37 157 IG-LIKE V-TYPE.

| | | | |
|--------------|----------------------------------|------|--|
| FT DOMAIN | 158 | 253 | LINK 1. |
| FT DOMAIN | 259 | 355 | LINK 2. |
| FT DOMAIN | 949 | 985 | EGF-LIKE 1. |
| FT DOMAIN | 987 | 1023 | EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL). |
| FT DOMAIN | 1025 | 1154 | C-TYPE LECTIN. |
| FT DOMAIN | 1155 | 1213 | SUSHI. |
| FT DISULFID | 58 | 139 | BY SIMILARITY. |
| FT DISULFID | 181 | 252 | BY SIMILARITY. |
| FT DISULFID | 205 | 226 | BY SIMILARITY. |
| FT DISULFID | 279 | 354 | BY SIMILARITY. |
| FT DISULFID | 303 | 324 | BY SIMILARITY. |
| FT DISULFID | 953 | 964 | BY SIMILARITY. |
| FT DISULFID | 958 | 973 | BY SIMILARITY. |
| FT DISULFID | 975 | 984 | BY SIMILARITY. |
| FT DISULFID | 1029 | 1040 | BY SIMILARITY. |
| FT DISULFID | 1057 | 1149 | BY SIMILARITY. |
| FT DISULFID | 1125 | 1141 | BY SIMILARITY. |
| FT DISULFID | 1156 | 1199 | BY SIMILARITY. |
| FT DISULFID | 1185 | 1212 | BY SIMILARITY. |
| FT CARBOHYD | 121 | 121 | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT CARBOHYD | 339 | 339 | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT CARBOHYD | 737 | 737 | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT CARBOHYD | 944 | 944 | O-LINKED (XYL. . .) (CHONDROITIN SULFATE). |
| FT CARBOHYD | 967 | 967 | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| FT CARBOHYD | 1164 | 1164 | N-LINKED (GLCNAC. . .) (POTENTIAL). |
| SEQ SEQUENCE | 1257 AA; 992B33DCFA19EE1B CRC64; | | |

Query Match 8.7%; Score 174; DB 1; Length 1257;
Best Local Similarity 30.8%; Pred. No. 1.8e-06;
Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;

| | |
|---------|---|
| QY 38 | QRCYKVIYFHDTRRLNFBEAKRRDGGQGVLSIEDKLEKFIENLLPSDGFV 97 |
| DB 1037 | QGHYR--YF---AHRRAWDAERDCRRRAGHLTSVHSPEBKFINF-----GHNSW 1085 |
| QY 98 | IGLRREEKQSNSTACQDLVAWTGSIQFBNMYVDEPS---CGSEVCVMVHQPAPAG 154 |
| DB 1086 | IGLNDRTVERD-----FQNTDNTGLQYENWREKQDNFAGDCVVMVHANG--- 1134 |
| QY 155 | IGGPFYMFQWDDRCNMKNFICK 177 |
| DB 1135 | -----RWNDVPCVNYLPYVCK 1150 |

RESULT 8
PGCV RAT
ID PGCV RAT STANDARD; PRT: 2738 AA.
AC Q9ERB4; O08592; O88564; Q9R1K4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE Chondroitin sulfate proteoglycan core protein 2 (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP) (Fragments).
GN CP5G2.
OS Rattus norvegicus (Rat).
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RX [1]
RN SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.
RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RT Wright T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
PL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [2]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=98308094; PubMed=9642104;
RA Milev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,

KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 20
 FT CHAIN 21 3358
 FT DOMAIN 21 146
 FT DOMAIN 167 244
 FT DOMAIN 265 346
 FT DOMAIN 348 1308
 FT DOMAIN 1309 3052
 FT DOMAIN 3052 3088
 FT DOMAIN 3090 3126
 FT DOMAIN 3139 3253
 FT DOMAIN 3258 3316
 FT DISULFID 44 130
 FT DISULFID 172 243
 FT DISULFID 196 217
 FT DISULFID 270 333
 FT DISULFID 294 315
 FT DISULFID 3056 3067
 FT DISULFID 3061 3076
 FT DISULFID 3078 3087
 FT DISULFID 3094 3105
 FT DISULFID 3099 3114
 FT DISULFID 3116 3125
 FT DISULFID 3132 3143
 FT DISULFID 3160 3252
 FT DISULFID 3228 3244
 FT DISULFID 3259 3302
 FT DISULFID 3288 3315
 FT CARBOHYD 57 57
 FT CARBOHYD 330 330
 FT CARBOHYD 351 351
 FT CARBOHYD 441 441
 FT CARBOHYD 807 807
 FT CARBOHYD 914 914
 FT CARBOHYD 951 951
 FT CARBOHYD 1305 1305
 FT CARBOHYD 1372 1372
 FT CARBOHYD 1679 1679
 FT CARBOHYD 2054 2054
 FT CARBOHYD 2244 2244
 FT CARBOHYD 2362 2362
 FT CARBOHYD 2627 2627
 FT CARBOHYD 3030 3030
 FT CARBOHYD 3332 3332
 FT CARBOHYD 3342 3342
 FT VARSPLIC 348 348
 FT VARSPLIC 349 1308
 FT VARSPLIC 1309 3052
 FT VARSPLIC 349 3052
 FT CONFLICT 126 126
 FT CONFLICT 348 348
 FT CONFLICT 1658 1658
 FT CONFLICT 1674 1680
 FT CONFLICT 3358 AA; 366938 MW; 071B80026BC0762D CRC64;
 SQ SEQUENCE 3358 AA; 366938 MW; 071B80026BC0762D CRC64;
 Query Match 8.7%; Score 174; DB 1; Length 3358;
 Best Local Similarity 28.5%; Pred. No. 6.8e-06;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
 QY 38 QPCQKVIYFHTSRRLNFEAKACRRDGGQLVSEDEQKLEFIENLPSDGP- 96
 DB 3140 QGQCYK---YF---AHRRTWDAARECRLOQAHLTSLSHSEQMPVNRV-----GHDYQ 3187
 QY 97 WIGLRREKQSNSTACQDLYAWTDGSGISQFRNMYVDEP-----SCGSEVCVMVHQSAP 152
 DB 3188 WIGL-----NDKMEFHDFRWTDGSLQYENWRPNQPSFFSAGDCVVIWHENG-- 3237
 QY 153 AGIGGPMYFOWNDRCNMKNFKICKYS-----DEKPAVPSREAGE 193

Db 3238 -----QNDVPCNVHLTYTCKGTACQPPVVENAKTFGK 3273
 RESULT 10
 ID_PGCV_HUMAN
 AC P13611; P20754; Q13010; Q13189; Q15123; Q9UNW5;
 DT 01-JAN-1990 (Rel. 13, created)
 DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 15-SEP-2003 (Rel. 42, last annotation update)
 DE Versican core protein precursor (large fibroblast proteoglycan)
 DE Chondroitin sulfate proteoglycan core protein 2 (PG-M) (Glia)
 DE hyaluronate-binding protein) (GHAP).
 GN CSPG2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM V0).
 RX MEDLINE=95105188; PubMed=7528742;
 RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
 RT "Characterization of the complete genomic structure of the human
 RT versican gene and functional analysis of its promoter.";
 RL J. Biol. Chem. 269:32999-33008(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM V1).
 RC TISSUE=Placenta;
 RX MEDLINE=90059882; PubMed=2583089;
 RA Zimmermann D.R., Ruoslahti E.;
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";
 RL EMBO J. 8:2975-2981(1989).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM V2).
 RC TISSUE=Glial tumor;
 RX MEDLINE=95105187; PubMed=7806529;
 RA Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "A novel glycosaminoglycan attachment domain identified in two
 RT alternative splice variants of human versican.";
 RL J. Biol. Chem. 269:32992-32998(1994).
 RN [4]
 RP SEQUENCE OF 2711-3396 FROM N.A.
 RC TISSUE=Lung fibroblast;
 RX MEDLINE=86007514; PubMed=2820964;
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
 RT lectin-like and growth factor-like sequences.";
 RL J. Biol. Chem. 262:13120-13125(1987).
 RN [5]
 RP SEQUENCE OF 251-347 FROM N.A.
 RX MEDLINE=93122792; PubMed=1478664;
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
 RA McPherson J.D.;
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
 RT human chromosome 5 (5q12-5q14).";
 RL Genomics 14:845-851(1992).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM V3).
 RC TISSUE=Brain;
 RX MEDLINE=95181355; PubMed=7876137;
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
 RT without a chondroitin sulfate attachment in region in mouse and human
 RT tissues.";
 RL J. Biol. Chem. 270:3914-3918(1995).
 RN [7]
 RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
 RC TISSUE=Aortic smooth muscle;
 RX MEDLINE=99370753; PubMed=10397680;
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
 RA Wight T.N.;
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";

RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
 RN [6]
 RP PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=89174663; PubMed=2466833;
 RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
 RT "Isolation and partial characterization of a glial
 RT hyaluronate-binding protein.";
 RL J. Biol. Chem. 264:5981-5987(1989).
 RN [9]
 RP TISSUE SPECIFICITY OF ISOFORMS.
 RX MEDLINE=96213482; PubMed=8627343;
 RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "Differential expression of versican isoforms in brain tumors.";
 RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
 CC -1- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -1- SUBUNIT: Interacts with PALN1 (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Comment=Additional isoforms seem to exist;
 CC Name=V0;
 CC IsoId=P13611-1; Sequence=Displayed;
 CC Name=V1;
 CC IsoId=P13611-2; Sequence=VSP_003082, VSP_003083;
 CC Name=V2;
 CC IsoId=P13611-3; Sequence=VSP_003084;
 CC Name=V3;
 CC IsoId=P13611-4; Sequence=VSP_003082, VSP_003085;
 CC Name=Vint;
 CC IsoId=P13611-5; Sequence=VSP_003086;
 CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
 CC in normal brain, gliomas, medulloblastomas, schwannomas,
 CC neurofibromas, and meningiomas; V2 is restricted to normal brain
 CC and gliomas; V3 is found in all these tissues except
 CC medulloblastomas.
 CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -1- SIMILARITY: Contains 2 link domains.
 CC -1- SIMILARITY: Contains 2 EGF-like domains.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U16306; AAA55018.1; -;
 DR EMBL; X15998; CAA34128.1; -;
 DR EMBL; X15998; CAA34128.1; -;
 DR EMBL; S52488; AAB24878.1; -;
 DR EMBL; U26555; AAA67565.1; -;
 DR EMBL; D32039; BAA66801.1; -;
 DR EMBL; J02814; AAA36437.1; -;
 DR EMBL; AF084545; AAD48545.1; -;
 DR PIR; S06014; A60979.
 DR HSSP; P01132; 1EGF.
 DR Genew; HGNC:2464; CSPG2.
 DR MIM; 118661; -;
 DR GO; GO:0005578; C:extracellular matrix; TAS.
 DR GO; GO:0005204; F:chondroitin sulfate proteoglycan; TAS.
 DR GO; GO:0005540; F:hyaluronic acid binding activity; TAS.
 DR GO; GO:0008037; P:cell recognition; TAS.
 DR GO; GO:0007275; P:development; TAS.
 DR InterPro; IPR000152; Asx hydroxyl.
 DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR007110; Ig-Like.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_Ca; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX HYDROXYL; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 DR PROSITE; PSS0041; C-TYPE LECTIN_2; 1.
 DR PROSITE; PS00022; EGF_1; 2.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_Ca; 1.
 DR PROSITE; PS0835; IG_LIKE; 1.
 DR PROSITE; PS01241; LINK; 2.
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 20
 FT CHAIN 21 3396
 FT DOMAIN 21 146
 FT DOMAIN 167 244
 FT DOMAIN 265 346
 FT DOMAIN 348 1335
 FT DOMAIN 1336 3089
 FT DOMAIN 3089 3125
 FT DOMAIN 3127 3163
 FT DOMAIN 3176 3290
 FT DOMAIN 3295 3353
 FT DISULFID 44 130
 FT DISULFID 172 243
 FT DISULFID 196 217
 FT DISULFID 270 345
 FT DISULFID 294 315
 FT DISULFID 3093 3104
 FT DISULFID 3098 3113
 FT DISULFID 3115 3124
 FT DISULFID 3131 3142
 FT DISULFID 3136 3151
 FT DISULFID 3153 3162
 FT DISULFID 3169 3180
 FT DISULFID 3197 3289
 FT DISULFID 3265 3281
 FT DISULFID 3296 3339
 FT DISULFID 3325 3352
 FT CARBOHYD 57 57
 FT CARBOHYD 330 330
 FT CARBOHYD 615 615
 Query Match 8.7%; Score 174; DB 1; Length 3396;
 Best Local Similarity 28.5%; Pred. No. 6.9e-06;
 Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
 QY 38 QRPCKYVYFHDTSRRLNFEAKEACRRDGGQVLSIESEDEQKLISKFIENLLPSDGF- 96
 Db 3177 QGQCYK-YF--AHRRTWDAERECRLQGAHLTSLSHHEQMFVNRV-----GHDIQ 3224
 QY 97 WTGLRRREKQSNSTACQDLYAWTDGSIQFRNYYDEP-----SCGSEVCVMYHPSAP 152
 Db 3225 WIGL-----NDKMFHDFRWTDGSLQYENWRNPQDPSPFSAGEDCVIWHENG-- 3274

QY 153 AGIGPYMFQWDRNCNNNNFICKYS-----DEKPAVPSREAGE 193
 Db 3275 -----QWNVPCVHLYTCKGTGTVACQGPVVENAKTFGK 3310

RESULT 11
 CD93_RAT STANDARD; PRT; 643 AA.
 AC Q9ET61; Q9J126;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Complement component, C1q receptor precursor (Complement component 1, q subcomponent, receptor 1) (C1qR) (C1qR(p)) (C1q/MBL/SPA receptor)
 DE (CD93 antigen) (Cell surface antigen AA4).
 GN C1QR1 OR CD93 OR C1QRP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PVG; TISSUE=Natural killer cells;
 RX MEDLINE=20545218; PubMed=11093153;
 RA Lovik G., Vaage J.T., Disen E., Szpirer C., Ryan J.C., Rolstad B.;
 RT "Characterization and molecular cloning of rat C1qR, a receptor on NK cells.";
 RL Eur. J. Immunol. 30:3355-3362 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wiistar; TISSUE=Lung;
 RX MEDLINE=20507883; PubMed=10934210;
 RA Dean Y.D., McGreal E.P., Akatsu H., Gaeque P.;
 RT "Molecular and cellular properties of the rat AA4 antigen, a C-type lectin-like receptor with structural homology to thrombomodulin.";
 RL J. Biol. Chem. 275:34382-34392 (2000).
 CC -I- FUNCTION: Receptor (or element of a larger receptor complex) for C1q, mannose-binding lectin (MBL2) and pulmonary surfactant monocytes and macrophages upon interaction with soluble defense collagens. May play a role in intercellular adhesion.
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -I- TISSUE SPECIFICITY: Widely expressed. Highly expressed in lung and heart. Expressed at lower level in brain, thymus, liver, spleen, intestine, kidney, adrenal gland, muscle and testis. Expressed on endothelial cells, platelets, undifferentiated monocytes and circulating natural killer cells.
 CC -I- PTM: N- and O-glycosylated (By similarity).
 CC -I- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -I- SIMILARITY: Contains 5 EGF-like domains.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF136537; AAG01572.1; -
 CC EMBL; AF160978; AAF80402.1; -
 CC HSSP; P35555; 1EMN
 CC GO; GO:0016021; C:integral to membrane; ISS.
 CC GO; GO:0004872; F:receptor activity; ISS.
 CC GO; GO:0016337; P:cell-cell adhesion; ISS.
 CC GO; GO:0042116; P:macrophage activation; ISS.
 CC GO; GO:0006909; P:phagocytosis; ISS.
 CC InterPro; IPR000152; Asx hydroxyl.
 CC InterPro; IPR001881; EGF_Ca.
 CC InterPro; IPR006209; EGF_Like.
 CC InterPro; IPR001304; Lectin_C.
 CC Pfam; PF00008; EGF; 5.
 CC Pfam; PF00059; lectin_c; 1.

Query Match 8.6%; Score 172.5; DB 1; Length 643;
 Best Local Similarity 25.9%; Pred. No. 9.8e-07;
 Matches 56; Conservative 37; Mismatches 74; Indels 49; Gaps 11;

QY 14 VLVGLRAATGRLLSGQP-----VCRGGTQRCYKVIYFHTSRRLNFEAKCRDQ 67
 Db 8 LLLGL-----LGQWAGAAADSEAVCEG-----TACYTAHW-----GKLSAAEAQHRCNENG 56
 QY 68 GQLVSISEDEQKLIBKFIENLL-----PSD-----GDFWIGLRRREKQSNSTACQDLVWT 120
 Db 57 GNLATVKSEEARHVEALQLLTKAPSETKIGKFWIGLQREKCKTYHDLPMKGFSWV 116
 QY 121 -DGSISQFRNWY-VDPSPGSEVGVVWY-----HQPSPAGIGGYPFMQWDDRC--- 168
 Db 117 GGGEDTYSYWNKASKSSCSIKRCVSLILDLSLKPSPHLP-----KWHESPGCTP 167
 QY 169 ----NMKNPFICKYSDKPAVSREAGEFETELTTP 200
 Db 168 DAPGNSIEGFLCKFPKMGCSPLALGGPGQLTYTTP 203

RESULT 12
 PGCV_CHICK
 ID PGCV_CHICK STANDARD; PRT; 3562 AA.
 AC Q90953; Q90945;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Versican core protein precursor (Large fibroblast proteoglycan) (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
 DE CP8G2.
 GN Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 3.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00615; C-TYPE_LECTIN_1; FALSE_NEG.
 DR PROSITE; PS00041; C-TYPE_LECTIN_2; 1.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01187; EGF_CA; 3.
 KW Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
 KW EGF-like domain; Lectin; Glycoprotein.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 643 COMPLEMENT COMPONENT C1Q RECEPTOR.
 FT DOMAIN 24 571 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 572 592 POTENTIAL.
 FT DOMAIN 593 643 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 31 173 C-TYPE LECTIN.
 FT DOMAIN 257 298 EGF-LIKE 1.
 FT DOMAIN 299 341 EGF-LIKE 2.
 FT DOMAIN 342 381 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 382 423 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 424 462 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DISULFID 261 272 BY SIMILARITY.
 FT DISULFID 268 282 BY SIMILARITY.
 FT DISULFID 284 297 BY SIMILARITY.
 FT DISULFID 303 314 BY SIMILARITY.
 FT DISULFID 308 325 BY SIMILARITY.
 FT DISULFID 327 340 BY SIMILARITY.
 FT DISULFID 346 355 BY SIMILARITY.
 FT DISULFID 351 364 BY SIMILARITY.
 FT DISULFID 366 380 BY SIMILARITY.
 FT DISULFID 386 397 BY SIMILARITY.
 FT DISULFID 393 406 BY SIMILARITY.
 FT DISULFID 408 422 BY SIMILARITY.
 FT DISULFID 428 437 BY SIMILARITY.
 FT DISULFID 433 446 BY SIMILARITY.
 FT DISULFID 448 461 BY SIMILARITY.
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 417 417 E -> K (IN REF. 2).
 SQ SEQUENCE 643 AA; 68781 MW; 9AE4C93AD943DB6 CRC64;

Db 3440 -----QWNVPCNYHLTYCKGTACGQPPVVENAKTFGK 3475

RESULT 13

```

ID _PGCA BOVIN STANDARD; PRT; 2364 AA.
AC P13608; P79117; Q28159;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core protein) (CSPCP).
DE AGC1.
GN Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Hering T.M., Kollar J., Huynh T.D.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 563-1056 FROM N.A.
RX MEDLINE=89380219; PubMed=2528543;
RA Anonson P., Heinegaard D., Oldberg A.;
RT "The keratan sulfate-enriched region of bovine cartilage proteoglycan consists of a consecutively repeated hexapeptide motif.";
RL J. Biol. Chem. 264:16170-16173(1989).
RN [3]
RP SEQUENCE OF 1609-2113 AND 2151-2364 FROM N.A.
RX MEDLINE=8727630; PubMed=3111460;
RA Oldberg A., Anonson P., Heinegaard D.;
RT "The partial amino acid sequence of bovine cartilage proteoglycan, deduced from a cDNA clone, contains numerous Ser-Gly sequences arranged in homologous repeats.";
RL Biochem. J. 243:255-259(1987).
RN [4]
RP SEQUENCE OF 2114-2150 FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93352525; PubMed=8349621;
RA Fuelop C., Walcz E., Valyon M., Glant T.T.;
RT "Expression of alternatively spliced epidermal growth factor-like domains in aggregates of different species. Evidence for a novel module.";
RL J. Biol. Chem. 268:17377-17383(1993).
RN [5]
RP PARTIAL SEQUENCE.
RX MEDLINE=8502710; PubMed=6489519;
RA Parin J.-P., Bonnet F., Jolles J., Jolles P.;
RT "Sequence data concerning the protein core of the cartilage proteoglycan monomers. Characterization of a sequence allowing the synthesis of an oligonucleotide probe.";
RL FEBS Lett. 176:37-42(1984).
RN [6]
RP PARTIAL SEQUENCE.
RX MEDLINE=87005253; PubMed=3530809;
RA Parin J.-P., Bonnet F., Jolles J.;
RT "Structural relationship between link proteins and proteoglycan monomers.";
RL FEBS Lett. 206:73-77(1986).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P13608-1; Sequence=Displayed;
CC Name=2;

```

```

CC IsoId=P13608-2; Sequence=VSP_003072;
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3, MAKES UP THE C-TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS) AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2 AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND O-LINKED (ABOUT 40) OLIGOSACCHARIDES.
CC -1- PTM: THE KERATAN SULFATE CONTENTS DIFFER CONSIDERABLY BETWEEN ADULT AND FETAL BOVINE PROTEOGLYCAN.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 4 link domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U76615; AAB38524.1; -
CC EMBL; L07053; -; NOT_ANNOTATED_CDS.
CC PIR; A34234; A39808.
CC PIR; T42630; T42630.
CC HSSP; P08709; 1BF9.
CC InterPro; IPR002353; AntifreezeII.
CC InterPro; IPR000152; Asx hydroxyl.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR003324; SGXSG.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00047; ig; 1.
CC Pfam; PF00059; lectin_C; 1.
CC Pfam; PF02339; SGXSG; 61.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 4.
CC PRINTS; PR01265; LINKMODULE.
CC PRINTS; PR00356; ANTIFREEZEII.
CC ProDom; PD000918; Link; 4.
CC SMART; SM00032; CCP; 1.
CC SMART; SM00034; CLECT; 1.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00445; LINK; 4.
CC PROSITE; PS00010; ASX_HYDROXYL; 1.
CC PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
CC PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01187; EGF_CA; 1.
CC PROSITE; PS00835; IG_LIKE; 1.
CC PROSITE; PS00290; IG_MHC; 1.
CC PROSITE; PS01241; LINK; 4.
CC Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
CC EGF-like domain; Calcium; Alternative splicing; Repeat;
CC Immunoglobulin domain.
CC SIGNAL 1 16 POTENTIAL.
CC CHAIN 17 2364 AGGRECAN CORE PROTEIN.
CC DOMAIN 25 147 IG-LIKE V-TYPE.
CC DOMAIN 170 247 LINK 1.
CC DOMAIN 268 349 LINK 2.
CC DOMAIN 504 581 LINK 3.

```

```

FT DOMAIN 602 683 LINK 4.
FT DOMAIN 774 907 23 X 6 AA APPROXIMATE TANDEM REPEATS OF
FT E-1EK]-P-P-P-S.
FT 1433 2112 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
FT 2113 2149 CS-2.
FT 2114 2364 G3.
FT 2161 2276 C-TYPE LECTIN.
FT 2280 2338 SUSHI.
FT 51 133 BY SIMILARITY.
FT DISULFID 175 246 BY SIMILARITY.
FT DISULFID 199 220 BY SIMILARITY.
FT DISULFID 273 348 BY SIMILARITY.
FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 509 580 BY SIMILARITY.
FT DISULFID 533 554 BY SIMILARITY.
FT DISULFID 607 682 BY SIMILARITY.
FT DISULFID 631 652 BY SIMILARITY.
FT DISULFID 2117 2128 BY SIMILARITY.
FT DISULFID 2182 2274 BY SIMILARITY.
FT DISULFID 2250 2266 BY SIMILARITY.
FT DISULFID 2281 2324 BY SIMILARITY.
FT DISULFID 2310 2337 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 2114 2150 Missing (in isoform 2).
FT SEQUENCE 2364 AA; 246359 MW; 6FF83763420C3D4C CRC64;
SQ
Query Match 7.9%; Score 158.5; DB 1; Length 2364;
Best Local Similarity 26.8%; Pred. No. 8.6e-05;
Matches 48; Conservative 29; Mismatches 59; Indels 43; Gaps 10;

QY 30 QPVCRG-...TQPCYKVIYFHTSRRLNFEAEACRDRGQGLVLSIESDEQKLIKFI 86
Db 2151 QKLCCEGTRFKQHCYR--HFPD---RATWDAESQCRKQSHLSVITPEEQ-----EFV 2201

QY 87 ENLLPSDGF-WGLRRREKQSNSTACQDIYAWTDSISQFRNYYVDEP-----SCSEV 141
Db 2202 NN---NAQYQWIGL-----NDKTIEDGRFMSDGHSLQFENWRPNQDNFFATGDC 2250

QY 142 CVVMYHOPSPAGIGPGYPMFQWDDRCNKNFNICKYS-----DEKPAVPSPREAGEETE 196
Db 2251 VVMWHEKG-----EMNDVPCNYQLPFTCKKGTACGEPVPPVHARIFGQKKD 2298

RESULT 14
PGCA HUMAN STANDARD; PRT; 2415 AA.
AC P16112; Q13650;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSPCp) (Chondroitin sulfate proteoglycan core protein 1).
GN AGC1 OR CSPG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORM 3).
RP TISSUE=Chondrocytes;
RC MEDLINE=91093289; PubMed=1985970;
RX Doege K.J., Sasaki M., Kimura T., Yamada Y.;
RA "Complete coding sequence and deduced primary structure of the human
RT cartilage large aggregating proteoglycan, aggrecan. Human-specific
RT repeats, and additional alternatively spliced forms.";
RL J. Biol. Chem. 266:1894-902(1991).
RN [2]

```

```

RP SEQUENCE OF 1778-2415 FROM N.A. (ISOFORM 2).
RC TISSUE=Chondrocytes;
RA Dudhia J., Hardingham T.E.;
RL Submitted (JAN-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RX MEDLINE=89380154; PubMed=2789216;
RA Baldwin C.T., Reginato A.M., Prockop D.J.;
RT "A new epidermal growth factor-like domain in the human core protein
RT for the large cartilage-specific proteoglycan. Evidence for
RT alternative splicing of the domain.";
RL J. Biol. Chem. 264:15747-15750(1989).
RN [4]
RP SEQUENCE OF 764-864 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=95128522; PubMed=7827755;
RA Barry F.P., Neame P.J., Sasse J., Pearson D.;
RT "Length variation in the keratan sulfate domain of mammalian
RT aggrecan.";
RL Matrix Biol. 14:323-328(1994).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION.
CC -1- SUBUNIT: Interacts with FBLN1 (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P16112-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P16112-2; Sequence=VSP_003074;
CC Name=3;
CC IsoId=P16112-3; Sequence=VSP_003074, VSP_003075;
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN. WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 4 link domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M55172; AAA62824.1; -
CC EMBL; J05062; AAA35726.1; -
CC EMBL; X17406; CAA35463.1; -
CC EMBL; S74659; AAC60643.2; -
CC PIR; A39086; A39086.
CC HSSP; P98066; 1TSG.
CC Genew; HGNC:319; AGC1.
CC MIM; 155760; -
CC GO; GO:0005204; P.chondroitin sulfate proteoglycan; TAS.
CC InterPro; IPR002353; Antifreeze1.
CC InterPro; IPR006209; EGF like.
CC InterPro; IPR006210; IEGF.
CC InterPro; IPR007110; Ig-like.

```

DR InterPro; IPR003599; IG.
 DR InterPro; IPR003006; IG.MHC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR003324; SGXSG.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 1.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00059; lectin_c; 1.
 DR Pfam; PF02339; SGXSG; 71.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 4.
 DR PRINTS; PR01265; LINKMODULE.
 DR PRINTS; PR00356; ANTIFREZEII.
 DR ProDom; PD000918; Link; 4.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 4.
 DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 DR PROSITE; PS00041; C-TYPE LECTIN_2; 1.
 DR PROSITE; PS00022; EGF 1; 1.
 DR PROSITE; PS01186; EGF 2; 1.
 DR PROSITE; PS00835; IG LIKE; 1.
 DR PROSITE; PS00290; IG.MHC; 1.
 DR PROSITE; PS01241; Link; 3.
 DR Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW EGF-like domain; Alternative splicing; Repeat; Immunoglobulin domain.
 FT SIGNAL 1 19
 FT CHAIN 20 2415
 FT DOMAIN 34 147
 FT DOMAIN 170 247
 FT DOMAIN 268 349
 FT DOMAIN 495 572
 FT DOMAIN 593 673
 FT DOMAIN 2164 2199
 FT DOMAIN 2201 2327
 FT DOMAIN 2331 2389
 FT DOMAIN 48 141
 FT DOMAIN 152 247
 FT DOMAIN 253 349
 FT DOMAIN 477 571
 FT DOMAIN 578 672
 FT DOMAIN 676 848
 FT DOMAIN 772 844
 FT DOMAIN 851 1497
 FT DOMAIN 941 1497
 FT DOMAIN 1498 2162
 FT DOMAIN 2163 2415
 FT DISULFID 51 133
 FT DISULFID 175 246
 FT DISULFID 199 220
 FT DISULFID 273 348
 FT DISULFID 297 318
 FT DISULFID 500 571
 FT DISULFID 524 545
 FT DISULFID 598 672
 FT DISULFID 621 642
 FT DISULFID 2168 2178
 FT DISULFID 2173 2187
 FT DISULFID 2189 2198
 FT DISULFID 2205 2216
 FT DISULFID 2233 2325
 FT DISULFID 2301 2317
 FT DISULFID 2332 2375
 FT DISULFID 2361 2388
 FT CARBOHYD 126 126
 FT CARBOHYD 239 239
 FT CARBOHYD 333 333
 FT CARBOHYD 387 387
 FT CARBOHYD 434 434
 FT CARBOHYD 602 602

FT CARBOHYD 657 657 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1898 1898 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 2163 2200 Missing (in isoform 2 and isoform 3).
 FT VARSPLIC 2330 2390 /FTID=VSP 003074.
 FT VARSPLIC 2330 2390 Missing (in isoform 3).
 FT CONFLICT 766 766 /FTID=VSP 003075.
 FT CONFLICT 847 847 E -> A (IN REF. 4).
 FT CONFLICT 1928 1928 E -> V (IN REF. 4).
 FT CONFLICT 1964 1964 I -> A (IN REF. 2).
 FT CONFLICT 2070 2070 I -> V (IN REF. 2 AND 3).
 FT CONFLICT 2070 2070 P -> A (IN REF. 3).
 FT CONFLICT 2391 2391 A -> P (IN REF. 2 AND 3).
 SQ SEQUENCE 2415 AA; 250191 MW; 1288937E1B98C6B6 CRC64;
 Query Match 7.9%; Score 158.5; DB 1; Length 2415;
 Best Local Similarity 29.1%; Pred. No. 8.9e-05;
 Matches 52; Conservative 22; Mismatches 62; Indels 43; Gaps 11;
 QY 30 QPVCRRG---TORPCYKVIYFHDTSRRLNFEAKACRRDGGQLVSISEDEQKLIETI 86
 DB 2202 QEVCEGWNKYQGHYR--HFPD---RETWDAERCRHQSHLSIVTPEQ----EFV 2252
 QY 87 ENLLPSDGGP-WIGLRRREKQSNSTACODLYAVTDSISQPRNRYVDEPS---CGSEVC 142
 DB 2253 NN---NAQDYQWIGL-----NDRITIEGDFRWSGDGHPMQENWRPNQDNFFAAGDC 2301
 QY 143 VVM-YHQPSAPAGIGPYFMFOWNDRCNMKNFKICKY-----DEKPAVPSRAEGEETE 196
 DB 2302 VVMWHEKG-----EWNQVPCNYHLPTCKGTGACGEPVVEHARTFQKKD 2349
 RESULT 15
 CD93 HUMAN
 ID_CD93 HUMAN STANDARD; PRT; 652 AA.
 AC QSNPY3; O00274;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Complement component C1q receptor precursor (Complement component 1, q
 DE subcomponent, receptor 1) (C1qR) (C1qR(p)) (C1q/MBL/SPA receptor)
 DE (CD93 antigen) (CDw93).
 GN C1QR1 OR CD93.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCSI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=97199258; PubMed=9047234;
 RA Nepomuceno R.R., Henschen-Edman A.H., Burgess W.H., Tenner A.J.;
 RT "cDNA cloning and primary structure analysis of C1qR(p), the human
 C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro.";
 RL Immunity 6:119-129(1997).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANT ALA-318.
 RX MEDLINE=21640567; PubMed=11791389;
 RA Steinberger P., Szekeres A., Willie S., Stockl J., Selenko N.,
 RA Prager E., Staffler G., Madic O., Stockinger H., Knapp W.;
 RT "Identification of human CD93 as the phagocytic C1q receptor (C1qRp)
 by expression cloning";
 RL J. Leukoc. Biol. 71:133-140(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Babbage A.K., Baguley C.L.,
 RA Jones M., Stavrides G., Almeida J.P., Bates K.N., Bead L.M., Beare D.M.,
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,

RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahnam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.B., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leivaesalaho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McEay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symamore N.C., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20";
RL Nature 414:865-871(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hake S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Mazny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickens M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smalls D.B.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP [5]
RX MEDLINE=21990337; PubMed=11994479;
RA McGreal E.P., Ikwaki N., Akatsu H., Morgan B.P., Gasque P.;
RT "Human C1qrp is identical with CD93 and the mMI-11 antigen but does
RL not bind C1q.";
RN J. Immunol. 168:5222-5232(2002).
RP [6]
RX MEDLINE=99192777; PubMed=10092817;
RA Nepomuceno R.R., Ruiz S., Park M., Tenner A.J.;
RT "C1qrp is a heavily O-glycosylated cell surface protein involved in
RL the regulation of phagocytic activity.";
RN J. Immunol. 162:3583-3589(1999).
CC -!- FUNCTION: Receptor (or element of a larger receptor complex) for
CC C1q, mannose-binding lectin (MBL2) and pulmonary surfactant
CC protein A (SpA). May mediate the enhancement of phagocytosis in
CC monocytes and macrophages upon interaction with soluble defense
CC collagens. May play a role in intercellular adhesion.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Highly expressed in endothelial cells,
CC platelets, cells of myeloid origin, such as monocytes and
CC neutrophils. Not expressed in cells of lymphoid origin.
CC -!- PTM: N- and O-glycosylated.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 5 EGF-like domains.
CC -!- CAUTION: Has been sometimes referred to as a collectin receptor.
CC -!- CAUTION: According to Ref.5, C1q is not a ligand for C1QR1.
CC -!- DATABASE: NAME=PROW; NOTE=PROW 3:1-6(2001);
CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/467246456_g.htm".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-eib.ch).
CC -----
CC EMBL; U94333; AAB53110.1; --
CC EMBL; AL118508; CAC00597.1; --
CC EMBL; BC028075; AAB28075.1; --
CC HSSP; P35555; 1EMN.
CC Genew; HGNC:15855; C1QR1.
CC MIM; 120577; --
CC GO; GO:0016021; C: integral to membrane; IC.
CC GO; GO:0004872; F: receptor activity; NAS.
CC GO; GO:0016337; P: cell-cell adhesion; IDA.
CC GO; GO:0042116; P: macrophage activation; NAS.
CC GO; GO:0006909; P: phagocytosis; NAS.
CC InterPro; IPR00152; Asx hydroxyl.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF_like.
CC InterPro; IPR001304; Lectin_C.
CC Pfam; PF00008; EGF; 5.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF01108; Tissue_fac; 1.
CC SMART; SM00034; CLECT; 1.
CC SMART; SM00179; EGF_CA; 3.
CC PROSITE; PS00010; ASX HYDROXYL; 3.
CC PROSITE; PS00615; C TYPE LECTIN_1; FALSE_NEG.
CC PROSITE; PS50041; C TYPE LECTIN_2; 1.
CC PROSITE; PS01186; EGF_2; 3.
CC PROSITE; PS01187; EGF_CA; 3.
CC Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
KW EGF-like domain; Lectin; Glycoprotein; Polymorphism.
FT SIGNAL 1 21
FT CHAIN 22 652 COMPLEMENT COMPONENT C1Q RECEPTOR.
FT DOMAIN 24 580 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 581 601 POTENTIAL.
FT DOMAIN 602 652 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 32 174 C-TYPE LECTIN.
FT DOMAIN 260 301 EGF-LIKE 1.
FT DOMAIN 302 344 EGF-LIKE 2.
FT DOMAIN 345 384 EGF-LIKE 3. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 385 426 EGF-LIKE 4. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 427 468 EGF-LIKE 5. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 594 601 POLY-LEU.
FT DISULFID 264 275 BY SIMILARITY.
FT DISULFID 271 285 BY SIMILARITY.
FT DISULFID 287 300 BY SIMILARITY.
FT DISULFID 306 317 BY SIMILARITY.
FT DISULFID 311 328 BY SIMILARITY.
FT DISULFID 330 343 BY SIMILARITY.
FT DISULFID 349 358 BY SIMILARITY.
FT DISULFID 354 367 BY SIMILARITY.
FT DISULFID 369 383 BY SIMILARITY.
FT DISULFID 389 400 BY SIMILARITY.
FT DISULFID 396 409 BY SIMILARITY.
FT DISULFID 411 425 BY SIMILARITY.
FT DISULFID 431 443 BY SIMILARITY.
FT DISULFID 439 452 BY SIMILARITY.
FT DISULFID 454 467 BY SIMILARITY.
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 318 318 V -> A.
FT /FTID=VAR 013573.
FT T -> V (IN REF. 1; AA SEQUENCE).
FT C -> T (IN REF. 1; AA SEQUENCE).
FT TA -> RI (IN REF. 1; AA SEQUENCE).
FT S -> N (IN REF. 1).
FT G -> A (IN REF. 1; AA SEQUENCE).
FT S -> A (IN REF. 1; AA SEQUENCE).
FT R -> Q (IN REF. 1; AA SEQUENCE).
CC CONFLICT 22 22
CC CONFLICT 36 36
CC CONFLICT 38 39
CC CONFLICT 155 155
CC CONFLICT 186 186
CC CONFLICT 492 492
CC CONFLICT 496 496

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:06:34 ; Search time 18.1827 Seconds
(without alignments)
1978.090 Million cell updates/sec

Title: US-09-887-855-2

Perfect score: 2000

Sequence: 1 MRPGTALQAVLLAVLLVGLR.....PDQMGSKSGWVENIYGY 374

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: Pirl:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 183 | 9.2 | 1456 | 1 A36563 | mannose receptor p |
| 2 | 178.5 | 8.9 | 1455 | 1 A48925 | mannose receptor p |
| 3 | 177 | 8.8 | 1268 | 2 S52781 | neurocan - mouse |
| 4 | 174.5 | 8.7 | 1643 | 2 T14274 | versican precursor |
| 5 | 174.5 | 8.7 | 3381 | 2 T42389 | versican precursor |
| 6 | 174 | 8.7 | 1257 | 2 S28764 | neurocan precursor |
| 7 | 174 | 8.7 | 2397 | 1 A55535 | versican precursor |
| 8 | 174 | 8.7 | 2409 | 1 A60979 | versican precursor |
| 9 | 171 | 8.6 | 3562 | 2 A47171 | chondroitin sulfat |
| 10 | 170.5 | 8.5 | 1479 | 2 T42710 | mannose receptor, |
| 11 | 158.5 | 7.9 | 1340 | 2 A39808 | proteoglycan core |
| 12 | 158.5 | 7.9 | 2227 | 2 T42630 | aggreccan - bovine |
| 13 | 158.5 | 7.9 | 2315 | 1 A39086 | aggreccan precursor |
| 14 | 156 | 7.8 | 162 | 1 LNRC1 | lectin BRA3-1 prec |
| 15 | 154.5 | 7.7 | 612 | 2 B42755 | E-selectin precurs |
| 16 | 153.5 | 7.7 | 2124 | 2 A28452 | proteoglycan core |
| 17 | 152 | 7.6 | 912 | 2 A54423 | brevican precursor |
| 18 | 151 | 7.5 | 162 | 1 LNRC3 | lectin BRA3-2 prec |
| 19 | 150.5 | 7.5 | 459 | 2 T24425 | hypothetical prote |
| 20 | 150 | 7.5 | 330 | 2 T42656 | hypothetical prote |
| 21 | 149 | 7.4 | 321 | 1 LNHRU | brevican - human (|
| 22 | 148.5 | 7.4 | 152 | 2 JCA4690 | IGE Fc receptor II |
| 23 | 148.5 | 7.4 | 404 | 2 A46274 | coagulation factor |
| 24 | 148.5 | 7.4 | 2132 | 1 A55182 | HIV gp120-binding |
| 25 | 148 | 7.4 | 253 | 2 E89130 | aggreccan precursor |
| 26 | 147 | 7.3 | 883 | 2 S57653 | protein F52E1.2 [i |
| 27 | 146 | 7.3 | 2109 | 1 I50421 | brevican precursor |
| 28 | 145.5 | 7.3 | 742 | 2 JC7595 | aggreccan precursor |
| 29 | 145 | 7.2 | 883 | 2 S49126 | scavenger receptor |
| | | | | | brevican precursor |

| | | | | | |
|----|-------|-----|------|-----------|--------------------|
| 30 | 144.5 | 7.2 | 173 | 2 S10548 | lectin - barnacle |
| 31 | 144.5 | 7.2 | 372 | 2 S23936 | L-selectin precurs |
| 32 | 143.5 | 7.2 | 129 | 2 JC4329 | coagulation factor |
| 33 | 143.5 | 7.2 | 372 | 1 A32375 | L-selectin precurs |
| 34 | 143.5 | 7.2 | 463 | 2 T26655 | hypothetical prote |
| 35 | 143 | 7.1 | 321 | 2 T26152 | hypothetical prote |
| 36 | 142.5 | 7.1 | 131 | 2 JC5058 | bitiscetin alpha c |
| 37 | 142 | 7.1 | 1487 | 2 S48719 | phospholipase-A(2) |
| 38 | 141.5 | 7.1 | 165 | 2 A47148 | reg I, regneratin |
| 39 | 141.5 | 7.1 | 331 | 1 LNMSER | IGE Fc receptor, 1 |
| 40 | 140 | 7.0 | 370 | 2 S22124 | L-selectin precurs |
| 41 | 139 | 7.0 | 248 | 1 LNHRUP5 | pulmonary surfacta |
| 42 | 139 | 7.0 | 248 | 1 LNHRUP1 | pulmonary surfacta |
| 43 | 139 | 7.0 | 248 | 1 LNHRUP1 | pulmonary surfacta |
| 44 | 139 | 7.0 | 283 | 1 LNPHLS | lectin precursor - |
| 45 | 138.5 | 6.9 | 172 | 2 S32489 | lectin - Iberian r |

ALIGNMENTS

RESULT 1

A36563

mannose receptor precursor - human

C:Species: Homo sapiens (man)

C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999

C:Accession: A36563; A60926; A44255; B44255; C44255; D44255; E44255; F44255; G44255; H4

R:Taylor, M.E.; Conary, J.T.; Lennartz, M.R.; Stahl, P.D.; Drickamer, K.

J. Biol. Chem. 265, 12156-12162, 1990

A:Title: Primary structure of the mannose receptor contains multiple motifs resembling

A:Reference number: A36563; MUID:90324192; PMID:2373685

A:Accession: A36563

A:Molecule type: mRNA

A:Residues: 1-1456 <TAY>

A:Cross-references: GB:J05550; NID:gl8675; PIDN:AAA59868.1; PID:gl88676

A>Note: Parts of this sequence, including the amino end of the mature protein, were cor

R: Szekowitz, R.A.B.; Sastry, K.; Bailly, P.; Warner, A.

J. Exp. Med. 172, 1785-1794, 1990

A:Title: Molecular characterization of the human macrophage mannose receptor: demonstra

A:Reference number: A60926; MUID:91079783; PMID:2258707

A:Accession: A60926

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1333, 'T', 1335-1456 <EZE>

A:Cross-references: GB:X55635

A>Note: translation of the nucleotide sequence is incomplete

A>Note: In the authors' translation additional residues Pro-Glu-Ile are shown after 497

R:Kim, S.J.; Ruiz, N.; Bezouska, K.; Drickamer, K.

Genomics 14, 721-727, 1992

A:Title: Organization of the gene encoding the human macrophage mannose receptor (MRC1)

A:Reference number: A44255; MUID:93052405; PMID:1294118

A:Accession: A44255

A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tr

A:Molecule type: DNA

A:Residues: 155-233, 'KSAL', 238-283;346-428;492-569;631-714, 716-719;783-820, 'N', 822-865;

A>Note: sequence extracted from NCBI backbone (NCBIP:118415, NCBIP:118421, NCBIP:118428

C:Genetics:

A:Gene: GDB:MRC1

A:Cross-references: GDB:133759; OMIM:153618

A:Map position: 10p13-10p13

C:Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II r

C:Keywords: duplication; lectin; tandem repeat; transmembrane protein

F:1-18/Domain: signal sequence #status predicted <SIG>

F:168-209/Domain: fibronectin type II repeat homology <2P1>

F:223-340/Domain: C-type lectin homology <LCH1>

F:362-486/Domain: C-type lectin homology <LCH2>

F:945-1079/Domain: C-type lectin homology <LCH3>

Query Match 9.2%; Score 183; DB 1; Length 1456;

Best Local Similarity 25.1%; Pred. No. 1.2e-06;

Matches 53; Conservative 38; Mismatches 66; Indels 54; Gaps 8;

QY 42 YKVIYFHTDTRLNPEAKEACRDCGQLVSTESDEQKLEKFTIENLLPSDGPWIGLR 101

Qy 97 WIGLRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEP-----SCGSEVCVMYHOPSAP 152
||||| WIGL-----NDKMFEDHFRWTDGTLQYENWRPNQPSFFSTGDCVVIWHENG-- 1521
Db 1472 WIGL-----NDKMFEDHFRWTDGTLQYENWRPNQPSFFSTGDCVVIWHENG-- 1521

Qy 153 AGIGGPFYFQWDDRCNMKNFICKYS-----DEKPAVPSREAGE----- 193
Db 1522 -----QWMDVPCNYHLTYTCKGTACQDLYAWTDGSIQFRNMYVDEP-----SCGSEVCVMYHOPSAP 152

Qy 194 -----ETELTT-----PVL-----PEETQEDAKTKFKESREAAIN 224
Db 1573 KDGFIQRHLPTIRCLNGRWAMPKICLNPSAYQRTYSKKYFKNSSAKON 1623

RESULT 5
T42389
N;Alternate names: versican form V0 - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 05-May-2000
C;Accession: T42389
R;Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
J. Biol. Chem. 273, 15758-15764, 1998
A;Title: Versican V2 is a major extracellular matrix component of the mature bovine brain
A;Reference number: Z17954; MUID:98288320; PMID:9624174
A;Accession: T42389
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-3381 <SCH>
A;Cross-references: EMBL:AF060456; NID:g3253299; PID:g3253300; PIDN:AAC24358.1
C;Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lectin
F;1-20/Domain: chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-3381/Product: versican, splice form V0 #status predicted <NAT>
F;57,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2356,28

Query Match 8.7%; Score 174.5; DB 2; Length 3381;
Best Local Similarity 25.5%; Pred. No. 1.7e-05;
Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;

Qy 38 QRPCYKVIYFHDTSRRLNFEAEACRRDGGQLVSTESDEQKLEKFIENLPSDGF 96
Db 3162 QGQCYK--YF--AHRRTWDAARECLQGAHLSILSHEEQMFVNRV-----GHDIQ 3209

Qy 97 WIGLRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEP-----SCGSEVCVMYHOPSAP 152
Db 3210 WIGL-----NDKMFEDHFRWTDGTLQYENWRPNQPSFFSTGDCVVIWHENG-- 3259

Qy 153 AGIGGPFYFQWDDRCNMKNFICKYS-----DEKPAVPSREAGE----- 193
Db 3260 -----QWMDVPCNYHLTYTCKGTACQDLYAWTDGSIQFRNMYVDEP-----SCGSEVCVMYHOPSAP 152

Qy 194 -----ETELTT-----PVL-----PEETQEDAKTKFKESREAAIN 224
Db 3311 KDGFIQRHLPTIRCLNGRWAMPKICLNPSAYQRTYSKKYFKNSSAKON 3361

RESULT 6
S28764
neurocan precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
C;Accession: S28764
R;Rauch, U.; Karthikeyan, L.; Maurel, P.; Margolis, R.U.; Margolis, R.K.
J. Biol. Chem. 267, 19536-19547, 1992
A;Title: Cloning and primary structure of neurocan, a developmentally regulated, aggregated
A;Reference number: S28764; MUID:92406907; PMID:1326557
A;Accession: S28764
A;Molecule type: mRNA
A;Residues: 1-1257 <RAU>
A;Cross-references: EMBL:M97161; NID:g205649; PIDN:AAC37679.1; PID:g205650
C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; EGF
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein

F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-1257/Product: neurocan #status predicted <NAT>
F;176-353/Domain: link protein repeat homology <LNK1>
F;274-355/Domain: link protein repeat homology <LNK2>
F;364-366/Region: cell attachment (R-G-D) motif
F;953-984/Domain: EGF homology <EGF>
F;1029-1149/Domain: C-type lectin homology <LCH>
F;1156-1212/Domain: complement factor H repeat homology <FHD>
F;121,339,737,967,1164/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;372,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F;944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

Query Match 8.7%; Score 174; DB 2; Length 1257;
Best Local Similarity 30.8%; Pred. No. 5.4e-06;
Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;

Qy 38 QRPCYKVIYFHDTSRRLNFEAEACRRDGGQLVSTESDEQKLEKFIENLPSDGF 97
Db 1037 QGQCYK--YF--AHRRTWDAARECLQGAHLSILSHEEQMFVNRV-----GHENSW 1085

Qy 98 IGLRRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEP-----CGSEVCVMYHOPSAP 154
Db 1086 IGLNDRTVKRD-----FQWTDNTGLQYENWRPNQPSFFSTGDCVVIWHENG-- 1134

Qy 155 IGGPFYFQWDDRCNMKNFICK 177
Db 1135 -----RWMDVPCNYHLTYTCKGTACQDLYAWTDGSIQFRNMYVDEP-----SCGSEVCVMYHOPSAP 154

RESULT 7
A55535
N;Alternate names: versican precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A55535
R;Ito, K.; Shinomura, T.; Zako, M.; Ujita, M.; Kimata, K.
J. Biol. Chem. 270, 958-965, 1995
A;Title: Multiple forms of mouse PG-M, a large chondroitin sulfate proteoglycan generated
A;Reference number: A55535; MUID:95122551; PMID:7822336
A;Accession: A55535
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-2397 <RES>
A;Cross-references: GB:D16263; NID:g862460; PIDN:BAA03796.1; PID:g862461
C;Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EGF
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-1654/Domain: versican #status predicted <NAT>
F;167-244/Domain: link protein repeat homology <LNK1>
F;265-346/Domain: link protein repeat homology <LNK2>
F;2095-2126/Domain: EGF homology <EG1>
F;2133-2164/Domain: EGF homology <EG2>
F;2171-2291/Domain: C-type lectin homology <LCH>
F;2298-2354/Domain: complement factor H repeat homology <FHD>

Query Match 8.7%; Score 174; DB 1; Length 2397;
Best Local Similarity 28.5%; Pred. No. 1.2e-05;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

Qy 38 QRPCYKVIYFHDTSRRLNFEAEACRRDGGQLVSTESDEQKLEKFIENLPSDGF 96
Db 2179 QGQCYK--YF--AHRRTWDAARECLQGAHLSILSHEEQMFVNRV-----GHDIQ 2226

Qy 97 WIGLRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEP-----SCGSEVCVMYHOPSAP 152
Db 2227 WIGL-----NDKMFEDHFRWTDGTLQYENWRPNQPSFFSTGDCVVIWHENG-- 2276

Qy 153 AGIGGPFYFQWDDRCNMKNFICKYS-----DEKPAVPSREAGE 193
Db 2277 -----QWMDVPCNYHLTYTCKGTACQDLYAWTDGSIQFRNMYVDEP-----SCGSEVCVMYHOPSAP 152

RESULT 8
A60979
versican precursor - human
N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan
N:Contains: glial hyaluronate-binding protein
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: S06014; S43921; A60979; A30358; A29348; A45131; I54179
R:Zimmermann, D.R.; Ruoslahti, E.
EMBL J. B. 2975-2981, 1989
A:Title: Multiple domains of the large fibroblast proteoglycan, versican.
A:Reference number: S06014; MUID:90059882; PMID:2583089
A:Accession: S06014
A:Molecule type: mRNA
A:Residues: 1-2409 <XIM>
A:CROSS-references: GB:X15998; NID:g37662; PIDN:CAA34128.1; PID:g37663
R:Yao, L.Y.; Moody, C.; Schoenher, E.; Wight, T.N.; Sandell, L.J.
Matrix Biol. 14, 213-225, 1994
A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by
A:Reference number: S43921; MUID:95005762; PMID:7921538
A:Accession: S43921
A:Molecule type: mRNA
A:Residues: 208-440;1094-1385;1910-2246 <YAO>
R:Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.
Brain Res. Bull. 22, 67-70, 1989
A:Title: Structural similarity of hyaluronate binding proteins in brain and cartilage.
A:Reference number: A60979; MUID:8929983; PMID:2469524
A:Accession: A60979
A:Molecule type: protein
A:Residues: 171-210;289-303 <BIG>
R:Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.
J. Biol. Chem. 264, 5981-5987, 1989
A:Title: Isolation and partial characterization of a glial hyaluronate-binding protein.
A:Reference number: A30358; MUID:89174663; PMID:2466833
A:Accession: A30358
A:Molecule type: protein
A:Residues: 24-50;80-87, 'D', 89-119;128-155;167-218;229-359, 'IR';261-268;277-283, 'G', 285-
R:Krusius, T.; Gehlsen, K.R.; Ruoslahti, E.
J. Biol. Chem. 262, 13120-13125, 1987
A:Title: A fibroblast chondroitin sulfate proteoglycan core protein contains lectin-like
A:Reference number: A29348; MUID:88007514; PMID:2820964
A:Accession: A29348
A:Molecule type: mRNA
A:Residues: 1725, 'V', 1727-2409 <KRU>
A:CROSS-references: GB:J02814
R:Perides, G.; Rahemtulla, F.; Lane, W.S.; Asher, R.A.; Bignami, A.
J. Biol. Chem. 267, 23883-23887, 1992
A:Title: Isolation of a large aggregating proteoglycan from human brain.
A:Reference number: A45131; MUID:93054750; PMID:1429726
A:Contents: brain
A:Accession: A45131
A:Molecule type: protein
A:Residues: 21-22, 'X', 24-37 <PE2>
A:Experimental source: brain
A:Note: sequence extracted from NCBI backbone (NCBIP:118884)
R:Iozzo, R.V.; Naso, M.F.; Cannizzaro, L.A.; Wasmuth, J.J.; McPherson, J.D.
Genomics 14, 845-851, 1992
A:Title: Mapping of the versican proteoglycan gene (CSPG2) to the long arm of human chro
A:Reference number: I54179; MUID:93122792; PMID:1478664
A:Accession: I54179
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 251-347 <RES>
A:CROSS-references: GB:S52488; NID:g263313; PIDN:AAB24878.1; PID:g263314
C:Genetics:
A:Gene: GDB:CSPG2
A:CROSS-references: GDB:127873; OMIM:118661
A:Map position: 5q12-5q14
A:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EG
F:1-30/Domain: signal sequence #status predicted <SIG>
F:21-2409/Product: proteoglycan 24K core protein #status predicted <MAT>
F:167-244/Domain: link protein repeat homology <LNK1>

F:265-346/Domain: link protein repeat homology <LNK2>
F:559-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>
F:2106-2137/Domain: EGF homology <EG1>
F:2144-2175/Domain: EGF homology <EG2>
F:2182-2302/Domain: C-type lectin homology <LCH>
F:2309-2365/Domain: complement factor H repeat homology <FHD>
Query Match 8.7%; Score 174; DB 1; Length 2409;
Best Local Similarity 28.5%; Pred. No. 1.2e-05;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
QY 38 QRCYKVIYFHDTSRLNPFEEAKEACRRDGGOLVSIESEDEQKLIKFKFIENLLPSDGF- 96
DB 2190 QCCYK--YF---AHRRTWDAARECRLOGAHLTSLSHSEEQMFVNRV-----GHDYQ 2237
QY 97 WTGLRRREEKQSNSTACQDLVYATWDTGSGISQFRNWTYDEP-----SCSEYCVVMYHQSAP 152
DB 2238 WTGL-----NDKMFEDHFRWTDGSLQYENWRPNQDPSFSGEDCVVLIWHENG-- 2287
QY 153 AGIGGYPYFQWDDRCNMKNPFICKYS-----DEKPAVPSREAGE 193
DB 2288 -----QWMDVPCNHYLTYTCKKGTVACGQPPVVENAKTFGK 2323
RESULT 9
A47171
chondroitin sulfate proteoglycan PG-M core protein - chicken
C:Species: Gallus gallus (chicken)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A47171
R:Shinomura, T.; Nishida, Y.; Ito, K.; Kimata, K.
J. Biol. Chem. 268, 14461-14469, 1993
A:Title: cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan expressed durin
A:Reference number: A47171; MUID:93300846; PMID:8314802
A:Accession: A47171
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-3562 <SHI>
A:CROSS-references: GB:D13542; NID:g391643; PIDN:BAA02742.1; PID:g391644
A:Experimental source: stage 22-23 developing limb buds
A:Note: sequence extracted from NCBI backbone (NCBIN:134456, NCBIP:134457)
C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lecti
F:166-243/Domain: link protein repeat homology <LNK1>
F:264-345/Domain: link protein repeat homology <LNK2>
F:3258-3289/Domain: EGF homology <EGF1>
F:3296-3327/Domain: EGF homology <EGF>
F:3334-3454/Domain: C-type lectin homology <LCH>
F:3461-3517/Domain: complement factor H repeat homology <FHD>
Query Match 8.6%; Score 171; DB 2; Length 3562;
Best Local Similarity 28.5%; Pred. No. 3.5e-05;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
QY 38 QRCYKVIYFHDTSRLNPFEEAKEACRRDGGOLVSIESEDEQKLIKFKFIENLLPSDGF- 96
DB 3342 QCCYK--YF---AHRRTWDAARECRLOGAHLTSLSHSEEQMFVNRV-----GHDYQ 3389
QY 97 WTGLRRREEKQSNSTACQDLVYATWDTGSGISQFRNWTYDEP-----SCSEYCVVMYHQSAP 152
DB 3390 WTGL-----NDKMFEDHFRWTDGSLQYENWRPNQDPSFSGEDCVVLIWHENG-- 3439
QY 153 AGIGGYPYFQWDDRCNMKNPFICKYS-----DEKPAVPSREAGE 193
DB 3440 -----QWMDVPCNHYLTYTCKKGTVACGQPPVVENAKTFGK 3475
RESULT 10
T42710
mannose receptor, macrophage - mouse
N:Alternate names: lambda lectin; phospholipase A2 receptor
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jun-2000

```

RESULT 13
A39086
aggrecan precursor, cartilage long splice form [validated] - human
NNAlternate names: chondroitin sulfate proteoglycan 1; large aggregating proteoglycan;
NNContains: aggrecan cartilage short splice form
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence revision 01-Dec-2000 #text change 08-Dec-2000

```


C;Accession: A39086; S50206; A43919; S46659; S66389; S68646; S62786; A34226; B43919; C43919
 R;Doerge, K.J.; Sasaki, M.; Kimura, T.; Yamada, Y.
 J. Biol. Chem. 265, 894-902, 1991
 A;Title: Complete coding sequence and deduced primary structure of the human cartilage 1
 A;Reference number: A39086; MUID:91093289; PMID:1985970
 A;Accession: A39086
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-2162, 2201-2329, 'A', 2392-2415 <DOE>
 A;Cross-references: GB:W55172; NID:G178258; PIDN:AAA62024.1; PID:G178259
 R;Glumoff, V.; Savontaus, M.; Vehanen, J.; Vuorio, E.
 Biochim. Biophys. Acta 1219, 613-622, 1994
 A;Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues by
 A;Reference number: S50206; MUID:95035091; PMID:7524681
 A;Accession: S50206
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 350-497 <GLU>
 A;Cross-references: EMBL:X80278; NID:9516295
 A;Note: this translation is not annotated in GenBank entry HSAGGREC, release 113.0
 R;Sandy, J.D.; Flannery, C.R.; Neame, P.J.; Lohmander, L.S.
 J. Clin. Invest. 89, 1512-1516, 1992
 A;Title: The structure of aggrecan fragments in human synovial fluid. Evidence for the
 domain.
 A;Reference number: A43919; MUID:92235266; PMID:1569188
 A;Accession: A43919
 A;Molecule type: protein
 A;Residues: 361-370, 'X', 372-373, 393-399, 'X', 401-407, 'X', 409 <SAS>
 A;Cross-references: PIDN:AA22079.1; PID:G248844; PIDN:AA22077.1; PID:G248842; PIDN:AA2
 A;Experimental source: synovial fluid
 A;Note: sequences modified after extraction from NCBI backbone
 R;Barry, F.P.; Neame, P.J.; Saase, J.; Pearson, D.
 Matrix Biol. 14, 323-328, 1994
 A;Title: Length variation in the keratan sulfate domain of mammalian aggrecan.
 A;Reference number: I46998; MUID:95128522; PMID:7827755
 A;Accession: S46659
 A;Molecule type: DNA
 A;Residues: 764-765, 'V', 767-846, 'V', 848-862, 'X', 864 <BAR>
 A;Cross-references: EMBL:S74659; NID:9807127; PIDN:AA60643.1; PID:9807128
 A;Note: the authors translated the codon GAA for residue 803 as Ala and Cyt for residue
 R;Illic, M.Z.; Mok, M.T.; Williamson, O.D.; Campbell, M.A.; Hughes, C.E.; Handley, C.J.
 Arch. Biochem. Biophys. 322, 22-30, 1995
 A;Title: Catabolism of aggrecan by explant cultures of human articular cartilage in the
 A;Reference number: S66389; MUID:96004775; PMID:7574678
 A;Accession: S66389
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 17-23, 24, 'X', 26-27, 393-401, 402-403 <ILI>
 R;Fosag, A.J.; Last, K.; Knaeuper, V.; Murphy, G.; Neame, P.J.
 FEBS Lett. 380, 17-20, 1996
 A;Title: Degradation of cartilage aggrecan by collagenase-3 (MMP-13).
 A;Reference number: S68646; MUID:96181659; PMID:8603731
 A;Accession: S68646
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 'V', 404-405, 'XX', <FOS>
 R;Dudhia, J.; Davidson, C.M.; Wells, T.M.; Vynios, D.H.; Hardingham, T.E.; Bayliss, M.T.
 Biochem. J. 313, 933-940, 1996
 A;Title: Age-related changes in the content of the C-terminal region of aggrecan in huma
 A;Reference number: S62786; MUID:96190740; PMID:8611178
 A;Accession: S62786
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1778-1927, 'A', 1929-1963, 'V', 1965-2162, 2201-2415 <DUD>
 A;Cross-references: EMBL:X17406; NID:930248; PIDN:CAA35463.1; PID:930249
 R;Baldwin, C.T.; Resinato, A.M.; Prockop, D.J.
 J. Biol. Chem. 264, 15747-15750, 1989
 A;Title: A new epidermal growth factor-like domain in the human core protein for the lat
 A;Reference number: A34226; MUID:89380154; PMID:2789216
 A;Accession: A34226
 A;Molecule type: mRNA
 A;Residues: 1936-1963, 'V', 1965-2069, 'A', 2071-2415 <BAL>

A;Cross-references: GB:J05062; NID:G181167; PIDN:AAA35726.1; PID:G181168
 C;Genetics;
 A;Gene: GDB:AGC1; CSFG1; CSPGCP; MSK16
 A;Cross-references: GDB:127479; OMIM:155760
 A;Map position: 15q26-15q26
 C;Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology; F
 C;Keywords: alternative splicing; cartilage; chondroitin sulfate proteoglycan; extracel
 F;1-19/Domain: signal sequence #status predicted <SIG>
 F;20-2162/Domain: aggrecan cartilage long splice form #status predicted <WATL>
 F;20-2162/Domain: aggrecan cartilage short splice form #status predicted <WATL>
 F;20-2162/Domain: aggrecan cartilage short splice form #status predicted <WATL>
 F;44-135/Domain: immunoglobulin homology <IMM>
 F;170-247/Domain: link protein repeat homology <LNK1>
 F;268-349/Domain: link protein repeat homology <LNK2>
 F;495-572/Domain: link protein repeat homology <LNK3>
 F;593-673/Domain: link protein repeat homology <LNK4>
 F;677-861/Domain: keratan sulfate attachment #status predicted <KSA>
 F;864-1510/Domain: chondroitin sulfate attachment #status predicted <CS1>
 F;1511-2162/Domain: chondroitin sulfate attachment #status predicted <CS2>
 F;2168-2198/Domain: EGF homology <EGF>
 F;2205-2325/Domain: C-type lectin homology <LCH>
 F;2322-2388/Domain: complement factor H repeat homology <FHD>
 F;126,239,333,387,434,602,657,737,1898/Binding site: carbohydrate (Asn) (covalent) #sta
 F;371,376/Binding site: keratan sulfate (Thr) (covalent) #status predicted
 Query Match 7.98; Score 158.5; DB 1; Length 2415;
 Best Local Similarity 29.18; Pred. No. 0.00023;
 Matches 52; Conservative 22; Mismatches 62; Indels 43; Gaps 11;
 QY 30 QVCRGG---TORPCYKVIYFHDTSRLNPEEAKACRRDGGQGVLSIESDEQKLIFKI 86
 DB 2202 QEVCEGWNKYQGHYR--HFPD---RETWDAERCRQOQSHLSIVTPEQ----EFV 2252
 QY 87 ENLLPDGDF-WGLRRREEKQSNSTACQDLYAWTQGISIQPRNYYVDEPS---CGSEVC 142
 DB 2253 NN---NAQDQWIGL-----NDRITGDFRWSGHPMQFENRPNQDNFFAAGDC 2301
 QY 143 VVM-VHQPAPAGIGGYPFMQWDDRCNMKNFICKYS-----DEKPAVPSREAGSETE 196
 DB 2302 VVMWHEKG-----EMNDVPCNVHLPTCKKGTVACGEPVVEHARTFGQKXD 2349
 RESULT 14
 LNRCL
 lectin BRA3-1 precursor - barnacle (Megabalanus rosa)
 C;Species: Megabalanus rosa
 C;Date: 24-Feb-1994 #sequence_revision 09-Sep-1994 #text_change 16-Jul-1999
 C;Accession: JCI503; A26094
 R;Takanatsu, N.; Takeda, T.; Kojima, M.; Heishi, M.; Muramoto, K.; Kamiya, H.; Shiba, T
 Gene 128, 251-255, 1993
 A;Title: Acorn barnacle Megabalanus rosa lectin (BRA-3): cDNA cloning, gene structure
 A;Reference number: JCI503; MUID:93292994; PMID:8514190
 A;Accession: JCI503
 A;Molecule type: DNA; mRNA
 A;Residues: 1-162 <TAK>
 A;Cross-references: DDBJ:D132299
 R;Muramoto, K.; Kamiya, H.
 Biochim. Biophys. Acta 874, 285-295, 1986
 A;Title: The amino-acid sequence of a lectin of the acorn barnacle Megabalanus rosa.
 A;Reference number: A26094
 A;Accession: A26094
 A;Molecule type: protein
 A;Residues: 25-145, 'K', 147-162 <MUR>
 A;Note: 146-Arg was also found
 C;Comment: This galactose-binding lectin is isolated from the coelomic fluid.
 C;Comment: This protein plays important roles in defense mechanisms and in development
 C;Comment: The molecule is a tetramer of identical chains.
 C;Genetics:
 A;Introns: 22/1; 47/2; 86/3
 C;Superfamily: tetranectin; C-type lectin homology
 C;Keywords: hemolymph; homotetramer; lectin
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-162/Product: lectin BRA3-1 #status experimental <MAT>

Search completed: December 22, 2003, 16:14:40
Job time : 20.1827 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:07:10 ; Search time 35.7385 Seconds
(without alignments)
1954.412 Million cell updates/sec

Title: US-09-887-855-2

Perfect score: 2000

Sequence: 1 MRPGTALQAVLLAVLLVGLR.....PDQGRSKESGWVENEIVGY 374

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
- 2: /cgn2_6/ptodata/2/pubpaa/ECT_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/2/pubpaa/ECTUS_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pap.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pap.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09A_NEW_PUB.pap.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pap.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pap.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pap.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 2000 | 100.0 | 374 | 9 | US-09-887-855-2 |
| 2 | 2000 | 100.0 | 374 | 12 | US-10-094-749-2090 |
| 3 | 2000 | 100.0 | 374 | 15 | US-10-149-819-15 |
| 4 | 1996 | 99.8 | 374 | 12 | US-10-094-749-2142 |
| 5 | 1986 | 99.3 | 382 | 10 | US-09-909-320-137 |
| 6 | 1986 | 99.3 | 382 | 10 | US-09-909-088B-137 |
| 7 | 1986 | 99.3 | 382 | 10 | US-09-905-291A-137 |
| 8 | 1986 | 99.3 | 382 | 10 | US-09-902-853-137 |
| 9 | 1986 | 99.3 | 382 | 10 | US-09-907-824-137 |
| 10 | 1986 | 99.3 | 382 | 10 | US-09-907-841-137 |
| 11 | 1986 | 99.3 | 382 | 11 | US-09-904-011-137 |
| 12 | 1986 | 99.3 | 382 | 11 | US-09-906-742-137 |
| 13 | 1986 | 99.3 | 382 | 11 | US-09-906-838-137 |
| 14 | 1986 | 99.3 | 382 | 11 | US-09-907-613-137 |
| 15 | 1986 | 99.3 | 382 | 11 | US-09-907-942-137 |

| | | | | | | |
|----|------|------|-----|----|--------------------|-------------------|
| 16 | 1986 | 99.3 | 382 | 11 | US-09-904-859-137 | Sequence 137, App |
| 17 | 1986 | 99.3 | 382 | 11 | US-09-909-204-137 | Sequence 137, App |
| 18 | 1986 | 99.3 | 382 | 11 | US-09-904-820-137 | Sequence 137, App |
| 19 | 1986 | 99.3 | 382 | 11 | US-09-904-786-137 | Sequence 137, App |
| 20 | 1986 | 99.3 | 382 | 11 | US-09-906-646-137 | Sequence 137, App |
| 21 | 1986 | 99.3 | 382 | 11 | US-09-906-786-137 | Sequence 137, App |
| 22 | 1986 | 99.3 | 382 | 11 | US-09-903-706-137 | Sequence 137, App |
| 23 | 1986 | 99.3 | 382 | 11 | US-09-902-903-137 | Sequence 137, App |
| 24 | 1986 | 99.3 | 382 | 11 | US-09-903-749A-137 | Sequence 137, App |
| 25 | 1986 | 99.3 | 382 | 11 | US-09-904-119-137 | Sequence 137, App |
| 26 | 1986 | 99.3 | 382 | 11 | US-09-904-956-137 | Sequence 137, App |
| 27 | 1986 | 99.3 | 382 | 11 | US-09-902-736-137 | Sequence 137, App |
| 28 | 1986 | 99.3 | 382 | 11 | US-09-907-794-137 | Sequence 137, App |
| 29 | 1986 | 99.3 | 382 | 11 | US-09-903-943-137 | Sequence 137, App |
| 30 | 1986 | 99.3 | 382 | 11 | US-09-904-462-137 | Sequence 137, App |
| 31 | 1986 | 99.3 | 382 | 11 | US-09-907-925-137 | Sequence 137, App |
| 32 | 1986 | 99.3 | 382 | 11 | US-09-902-692-137 | Sequence 137, App |
| 33 | 1986 | 99.3 | 382 | 11 | US-09-903-520-137 | Sequence 137, App |
| 34 | 1986 | 99.3 | 382 | 11 | US-09-905-056-137 | Sequence 137, App |
| 35 | 1986 | 99.3 | 382 | 11 | US-09-909-064-137 | Sequence 137, App |
| 36 | 1986 | 99.3 | 382 | 11 | US-09-904-553-137 | Sequence 137, App |
| 37 | 1986 | 99.3 | 382 | 11 | US-09-905-381-137 | Sequence 137, App |
| 38 | 1986 | 99.3 | 382 | 11 | US-09-905-088-137 | Sequence 137, App |
| 39 | 1986 | 99.3 | 382 | 11 | US-09-907-575-137 | Sequence 137, App |
| 40 | 1986 | 99.3 | 382 | 11 | US-09-905-075-137 | Sequence 137, App |
| 41 | 1986 | 99.3 | 382 | 11 | US-09-902-759-137 | Sequence 137, App |
| 42 | 1986 | 99.3 | 382 | 11 | US-09-902-634-137 | Sequence 137, App |
| 43 | 1986 | 99.3 | 382 | 11 | US-09-902-713-137 | Sequence 137, App |
| 44 | 1986 | 99.3 | 382 | 11 | US-09-907-979-137 | Sequence 137, App |
| 45 | 1986 | 99.3 | 382 | 11 | US-09-902-615-137 | Sequence 137, App |

ALIGNMENTS

RESULT 1

US-09-887-855-2
; Sequence 2, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-887-855-2

Query Match 100.0%; Score 2000; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.3e-179; Indels 0; Gaps 0;
Matches 374; Conservative 0; Mismatches 0

| | | | |
|----|-----|---|-----|
| Qy | 1 | MRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRGGTQPCVKYIYFHTSRLNFEAK | 60 |
| Db | 1 | MRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRGGTQPCVKYIYFHTSRLNFEAK | 60 |
| Qy | 61 | EACRRDGGQLYSIESEDEOKLIEKFIENLLPSDGFNGLRRREKQSNSTACQDLYAWT | 120 |
| Db | 61 | EACRRDGGQLYSIESEDEOKLIEKFIENLLPSDGFNGLRRREKQSNSTACQDLYAWT | 120 |
| Qy | 121 | DGSI SQFNWYVDPSPCGSEVCVVMYHQPAPAGIGGYPYFMQNDRCNMKNPFCKYSD | 180 |
| Db | 121 | DGSI SQFNWYVDPSPCGSEVCVVMYHQPAPAGIGGYPYFMQNDRCNMKNPFCKYSD | 180 |
| Qy | 181 | EKPAVPSREAGGEETLTPVLPETQEDAKKTFKESREALNLAYILIPSIPLLLLV | 240 |
| Db | 181 | EKPAVPSREAGGEETLTPVLPETQEDAKKTFKESREALNLAYILIPSIPLLLLV | 240 |

Qy 241 VTTVVCWVICRKRKREQDPSTKKQHTIWPSPHQGNSPDLEVYVIRKQSEADLAETRP 300
Db 241 VTTVVCWVICRKRKREQDPSTKKQHTIWPSPHQGNSPDLEVYVIRKQSEADLAETRP 300
Qy 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYFSPDQNGR 360
Db 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYFSPDQNGR 360
Qy 361 SKESGWENEIYGY 374
Db 361 SKESGWENEIYGY 374

RESULT 2

US-10-094-749-2090
; Sequence 2090, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KETICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2090
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2090

Query Match 100.0%; Score 2000; DB 12; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.3e-179;
Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MRPGTALQAVLLAVLLVGLRAATGRLLSGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAK 60
Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAK 60
Qy 61 EACRRDGGQLVSESEDEQKLIKFIENLLPSDGFWMGLRRRREKQSNSTACQDLYAWT 120
Db 61 EACRRDGGQLVSESEDEQKLIKFIENLLPSDGFWMGLRRRREKQSNSTACQDLYAWT 120
Qy 121 DGSISQFRNMYVDEPSCGSEVCVVMYHQPAPAGIGGPFYQWDDRCNKNPFICKYSD 180
Db 121 DGSISQFRNMYVDEPSCGSEVCVVMYHQPAPAGIGGPFYQWDDRCNKNPFICKYSD 180
Qy 181 EKPAVPSREAGBETELTTPVLPEETQEDAKTKFKESREAAALNLAIVILPSIPLLLLV 240
Db 181 EKPAVPSREAGBETELTTPVLPEETQEDAKTKFKESREAAALNLAIVILPSIPLLLLV 240

Qy 241 VTTVVCWVICRKRKREQDPSTKKQHTIWPSPHQGNSPDLEVYVIRKQSEADLAETRP 300
Db 241 VTTVVCWVICRKRKREQDPSTKKQHTIWPSPHQGNSPDLEVYVIRKQSEADLAETRP 300
Qy 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYFSPDQNGR 360
Db 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYFSPDQNGR 360
Qy 361 SKESGWENEIYGY 374
Db 361 SKESGWENEIYGY 374

RESULT 3

US-10-149-819-15
; Sequence 15, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: SHAH, Purvi
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PF-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 3143411CD1
US-10-149-819-15

Query Match 100.0%; Score 2000; DB 15; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.3e-179;
Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRPGTALQAVLLAVLLVGLRAATGRLLSGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAK 60
Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSGQPCVCRGGTQPCYKVIYFHDTSRRLNFEAK 60
Qy 61 EACRRDGGQLVSESEDEQKLIKFIENLLPSDGFWMGLRRRREKQSNSTACQDLYAWT 120
Db 61 EACRRDGGQLVSESEDEQKLIKFIENLLPSDGFWMGLRRRREKQSNSTACQDLYAWT 120
Qy 121 DGSISQFRNMYVDEPSCGSEVCVVMYHQPAPAGIGGPFYQWDDRCNKNPFICKYSD 180
Db 121 DGSISQFRNMYVDEPSCGSEVCVVMYHQPAPAGIGGPFYQWDDRCNKNPFICKYSD 180
Qy 181 EKPAVPSREAGBETELTTPVLPEETQEDAKTKFKESREAAALNLAIVILPSIPLLLLV 240
Db 181 EKPAVPSREAGBETELTTPVLPEETQEDAKTKFKESREAAALNLAIVILPSIPLLLLV 240
Qy 241 VTTVVCWVICRKRKREQDPSTKKQHTIWPSPHQGNSPDLEVYVIRKQSEADLAETRP 300
Db 241 VTTVVCWVICRKRKREQDPSTKKQHTIWPSPHQGNSPDLEVYVIRKQSEADLAETRP 300
Qy 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYFSPDQNGR 360
Db 301 DLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYFSPDQNGR 360


```
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-320-137

Query Match      99.3%; Score 1986; DB 10; Length 382;
Best Local Similarity 97.9%; Pred. No. 2.7e-178;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGTQPCVKVIYFHDTSR 52
   |||||
Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLRLGGQPCVCRGGTQPCVKVIYFHDTSR 60
   |||||

QY 53 RLNFEEAKEACRRDGGQGVLSIESEDEQKLEKFIENLLPSDGFWMGLRREEKQSNSTA 112
   |||||
Db 61 RLNFEEAKEACRRDGGQGVLSIESEDEQKLEKFIENLLPSDGFWMGLRREEKQSNSTA 120
   |||||

QY 113 CODLYAWTDGSIQFRNMYVDEPSCGSEVCVMYHQPAPAGIGGPFYMFQWDDRCNMKN 172
   |||||
Db 121 CODLYAWTDGSIQFRNMYVDEPSCGSEVCVMYHQPAPAGIGGPFYMFQWDDRCNMKN 180
   |||||

QY 173 NFICKYSDEKPAVPSREAEGETELTTPVLPEETOEDAKKTFKESREAAALNLAYILIPS 232
   |||||
Db 181 NFICKYSDEKPAVPSREAEGETELTTPVLPEETOEDAKKTFKESREAAALNLAYILIPS 240
   |||||

QY 233 IPLLILLVVTVCVWVICRKRKEQDPDSTKKQHTIWPSPHQNSPDLEVNIVIRKOSE 292
   |||||
Db 241 IPLLILLVVTVCVWVICRKRKEQDPDSTKKQHTIWPSPHQNSPDLEVNIVIRKOSE 300
   |||||

QY 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNNVMPSSGGFVTLVSSESGFVTNDIYE 352
   |||||
Db 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNNVMPSSGGFVTLVSSESGFVTNDIYE 360
   |||||

QY 353 FSPDQMGSKESGWVENEIYGY 374
   |||||
Db 361 FSPDQMGSKESGWVENEIYGY 382
   |||||
```

RESULT 6

```
US-09-909-088B-137
; Sequence 137, Application US/09909088B
; Patent No. US20020146709A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Flivaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
```

```
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,088B
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-088B-137
```

```
Query Match      99.3%; Score 1986; DB 10; Length 382;
Best Local Similarity 97.9%; Pred. No. 2.7e-178;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGTQPCVKVIYFHDTSR 52
   |||||
Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLRLGGQPCVCRGGTQPCVKVIYFHDTSR 60
   |||||

QY 53 RLNFEEAKEACRRDGGQGVLSIESEDEQKLEKFIENLLPSDGFWMGLRREEKQSNSTA 112
   |||||
Db 61 RLNFEEAKEACRRDGGQGVLSIESEDEQKLEKFIENLLPSDGFWMGLRREEKQSNSTA 120
   |||||

QY 113 CODLYAWTDGSIQFRNMYVDEPSCGSEVCVMYHQPAPAGIGGPFYMFQWDDRCNMKN 172
   |||||
Db 121 CODLYAWTDGSIQFRNMYVDEPSCGSEVCVMYHQPAPAGIGGPFYMFQWDDRCNMKN 180
   |||||

QY 173 NFICKYSDEKPAVPSREAEGETELTTPVLPEETOEDAKKTFKESREAAALNLAYILIPS 232
   |||||
Db 181 NFICKYSDEKPAVPSREAEGETELTTPVLPEETOEDAKKTFKESREAAALNLAYILIPS 240
   |||||

QY 233 IPLLILLVVTVCVWVICRKRKEQDPDSTKKQHTIWPSPHQNSPDLEVNIVIRKOSE 292
   |||||
Db 241 IPLLILLVVTVCVWVICRKRKEQDPDSTKKQHTIWPSPHQNSPDLEVNIVIRKOSE 300
   |||||

QY 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNNVMPSSGGFVTLVSSESGFVTNDIYE 352
   |||||
Db 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNNVMPSSGGFVTLVSSESGFVTNDIYE 360
   |||||

QY 353 FSPDQMGSKESGWVENEIYGY 374
   |||||
```

Db 361 FSPQMGSKSGWVENEIYG 382
|||||
US-09-905-291A-137
RESULT 7
; Sequence 137, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gottard, A.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05

; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-291A-137
Query Match 99.3%; Score 1986; DB 10; Length 382;
Beat Local Similarity 97.9%; Pred. No. 2.7e-178;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
Qy 1 MRPTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGTQRPCYKVIYFHDTSR 52
Db 1 MRPTALQAVLLAVLLVGLRAATGRLLSASDLRLGGQPVCRGGTQRPCYKVIYFHDTSR 60
Qy 53 RLNFEEAKEACRRDGGQGLVSIESTEDEQKLEKFTENLLPSDGDFTWIGLRRREEKQSNSTA 112
Db 61 RLNFEEAKEACRRDGGQGLVSIESTEDEQKLEKFTENLLPSDGDFTWIGLRRREEKQSNSTA 120
Qy 113 QDLYAWTDGSIQFRNMYVDEPSCGSEVCVVMYTHQSPAPAGIGPYMFQNDRCNMKN 172
Db 121 QDLYAWTDGSIQFRNMYVDEPSCGSEVCVVMYTHQSPAPAGIGPYMFQNDRCNMKN 180
Qy 173 NFICKYSDEKPAVPSREABGESETLTPVLPETQEEDAKKTFKESREAAALNLAYILIPS 232
Db 181 NFICKYSDEKPAVPSREABGESETLTPVLPETQEEDAKKTFKESREAAALNLAYILIPS 240
Qy 233 IPLLILLAVTTVWCWVICRKRKEQDPSTKKOHTIWPSPHQNSPDLEVYVIRKOSE 292
Db 241 IPLLILLAVTTVWCWVICRKRKEQDPSTKKOHTIWPSPHQNSPDLEVYVIRKOSE 300
Qy 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFTVNDIYE 352
Db 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFTVNDIYE 360
Qy 353 FSPQMGSKSGWVENEIYG 374
Db 361 FSPQMGSKSGWVENEIYG 382

RESULT 8

US-09-902-853-137
; Sequence 137, Application US/09902853
; Publication No. US20020192659A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gottard, A.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14

;; CURRENT APPLICATION NUMBER: US/09/902,853
;; CURRENT FILING DATE: 2001-07-10
;; PRIOR APPLICATION NUMBER: US/09/665,350
;; PRIOR FILING DATE: 2000-09-18
;; PRIOR APPLICATION NUMBER: US 60/143,048
;; PRIOR FILING DATE: 1999-07-07
;; PRIOR APPLICATION NUMBER: US 60/145,698
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: US 60/146,222
;; PRIOR FILING DATE: 1999-07-28
;; PRIOR APPLICATION NUMBER: PCT/US99/20594
;; PRIOR FILING DATE: 1999-09-08
;; PRIOR APPLICATION NUMBER: PCT/US99/20944
;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/21547
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/23089
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: PCT/US99/28214
;; PRIOR FILING DATE: 1999-11-29
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: 1999-11-30
;; PRIOR APPLICATION NUMBER: PCT/US99/28564
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/28565
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: PCT/US99/30911
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US99/30999
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US00/00219
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 137
;; LENGTH: 382
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-902-853-137

Query Match 99.3%; Score 1986; DB 10; Length 382;
Best Local Similarity 97.9%; Pred. No. 2.7e-178;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 MRFGTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGQTQPCYKVIYFHDTSR 52
DB 1 MRFGTALQAVLLAVLLVGLRAATGRLLSASDLDLRGGQVCRGQTQPCYKVIYFHDTSR 60

QY 53 RLNFEAKEACRRDGGQVLSIESEDEOKLIEKIENLLPSDGFWGLRRREKQSNSTA 112
DB 61 RLNFEAKEACRRDGGQVLSIESEDEOKLIEKIENLLPSDGFWGLRRREKQSNSTA 120

QY 113 CQDLYAWTDGSIQFRNWWYDFPSCGSEVCVMYHQPAPAGIGGYPMFQWDDRCNMKN 172
DB 121 CQDLYAWTDGSIQFRNWWYDFPSCGSEVCVMYHQPAPAGIGGYPMFQWDDRCNMKN 180

QY 173 NFICKYSDEKPAVPAREAGEETELTPVLPESTQEDAKTKPKESREAAALNAYILIPS 232
DB 181 NFICKYSDEKPAVPAREAGEETELTPVLPESTQEDAKTKPKESREAAALNAYILIPS 240

QY 233 IPLLILLVTVVVCVWVICRKRKEQPDSTKKQHTIWPSPHOGNSPDLEVVNVIRKQSE 292
DB 241 IPLLILLVTVVVCVWVICRKRKEQPDSTKKQHTIWPSPHOGNSPDLEVVNVIRKQSE 300

QY 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFTNDIYE 352
DB 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFTNDIYE 360

QY 353 FSPDQMGSKESGWENEIYGY 374
|||||

Db 361 FSPDQMGSKESGWENEIYGY 382

RESULT 9
US-09-907-824-137
;; Sequence 137, Application US/09907824
;; Publication No. US20020197671A1
;; GENERAL INFORMATION:
;; APPLICANT: Genentech, Inc.
;; APPLICANT: Ashkenazi, Avi
;; APPLICANT: Botstein, David
;; APPLICANT: Deenoyers, Luc
;; APPLICANT: Eaton, Dan L.
;; APPLICANT: Ferrara, Napoleone
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Fong, Sherman
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerber, Hanspeter
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, A.
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Grimaldi, Christopher J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Hillan, Kenneth, J.
;; APPLICANT: Kljavin, Ivar J.
;; APPLICANT: Mather, Jennie P.
;; APPLICANT: Pan, James
;; APPLICANT: Paoni, Nicholas F.
;; APPLICANT: Roy, Margaret Ann
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Williams, P. Mickey
;; APPLICANT: Wood, William, I.
;; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;; FILE REFERENCE: 10466-14
;; CURRENT APPLICATION NUMBER: US/09/907,824
;; CURRENT FILING DATE: 2001-07-17
;; PRIOR APPLICATION NUMBER: 09/665,350
;; PRIOR FILING DATE: 2000-09-18
;; PRIOR APPLICATION NUMBER: PCT/US00/04414
;; PRIOR FILING DATE: 2000-02-22
;; PRIOR APPLICATION NUMBER: US 60/143,048
;; PRIOR FILING DATE: 1999-07-07
;; PRIOR APPLICATION NUMBER: US 60/145,698
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: US 60/146,222
;; PRIOR FILING DATE: 1999-07-28
;; PRIOR APPLICATION NUMBER: PCT/US99/20594
;; PRIOR FILING DATE: 1999-09-08
;; PRIOR APPLICATION NUMBER: PCT/US99/20944
;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/21547
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/23089
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: PCT/US99/28214
;; PRIOR FILING DATE: 1999-11-29
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: 1999-11-30
;; PRIOR APPLICATION NUMBER: PCT/US99/28564
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/28565
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: PCT/US99/30911
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US99/30999
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US00/00219

;; PRIOR FILING DATE: 2000-01-05
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 137
;; LENGTH: 382
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-907-824-137

Query Match
Best Local Similarity 99.3%; Score 1986; DB 10; Length 382;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 MRPGTALQVLLAVLLVGLRAATGRLLS-----GQVCRGGTQPCYKVIYFHDTSR 52
DB 1 MRPGTALQVLLAVLLVGLRAATGRLLSASDLRLGGQPVCRGGTQPCYKVIYFHDTSR 60
QY 53 RLNFEEAKACRRDGGQLVSISEDEOKLIEKFIENLLPSDGFWIGLRRREKQSNSTA 112
DB 61 RLNFEEAKACRRDGGQLVSISEDEOKLIEKFIENLLPSDGFWIGLRRREKQSNSTA 120
QY 113 CODLYAWTDGSIQFRNMYVDEPSCGSEVCVVMYHQPSPAPAGIGGPFQWDDRCNMKN 172
DB 121 CODLYAWTDGSIQFRNMYVDEPSCGSEVCVVMYHQPSPAPAGIGGPFQWDDRCNMKN 180
QY 173 NFICKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNLYILIPS 232
DB 181 NFICKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNLYILIPS 240
QY 233 IPLLALLVVTTCVVMWICRKRKEQPDPTTKKQHTIWPSPHQGNSPDLVYVNRKQSE 292
DB 241 IPLLALLVVTTCVVMWICRKRKEQPDPTTKKQHTIWPSPHQGNSPDLVYVNRKQSE 300
QY 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFTVNDIYE 352
DB 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFTVNDIYE 360
QY 353 FSPDQGRSKESGWVENEIYGY 374
DB 361 FSPDQGRSKESGWVENEIYGY 382

RESULT 10

US-09-907-841-137
; Sequence 137, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same

;; FILE REFERENCE: 10466-14
;; CURRENT APPLICATION NUMBER: US/09/907,841
;; CURRENT FILING DATE: 2001-11-20
;; PRIOR APPLICATION NUMBER: PCT/US00/04414
;; PRIOR FILING DATE: 2000-02-22
;; PRIOR APPLICATION NUMBER: US 60/143,048
;; PRIOR FILING DATE: 1999-07-07
;; PRIOR APPLICATION NUMBER: US 60/145,698
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: US 60/146,222
;; PRIOR FILING DATE: 1999-07-28
;; PRIOR APPLICATION NUMBER: PCT/US99/20594
;; PRIOR FILING DATE: 1999-09-08
;; PRIOR APPLICATION NUMBER: PCT/US99/20944
;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/21547
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/23089
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: PCT/US99/28214
;; PRIOR FILING DATE: 1999-11-29
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 137
;; LENGTH: 382
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-907-841-137

Query Match
Best Local Similarity 99.3%; Score 1986; DB 10; Length 382;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 MRPGTALQVLLAVLLVGLRAATGRLLS-----GQVCRGGTQPCYKVIYFHDTSR 52
DB 1 MRPGTALQVLLAVLLVGLRAATGRLLSASDLRLGGQPVCRGGTQPCYKVIYFHDTSR 60
QY 53 RLNFEEAKACRRDGGQLVSISEDEOKLIEKFIENLLPSDGFWIGLRRREKQSNSTA 112
DB 61 RLNFEEAKACRRDGGQLVSISEDEOKLIEKFIENLLPSDGFWIGLRRREKQSNSTA 120
QY 113 CODLYAWTDGSIQFRNMYVDEPSCGSEVCVVMYHQPSPAPAGIGGPFQWDDRCNMKN 172
DB 121 CODLYAWTDGSIQFRNMYVDEPSCGSEVCVVMYHQPSPAPAGIGGPFQWDDRCNMKN 180
QY 173 NFICKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNLYILIPS 232
DB 181 NFICKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNLYILIPS 240
QY 233 IPLLALLVVTTCVVMWICRKRKEQPDPTTKKQHTIWPSPHQGNSPDLVYVNRKQSE 292
DB 241 IPLLALLVVTTCVVMWICRKRKEQPDPTTKKQHTIWPSPHQGNSPDLVYVNRKQSE 300
QY 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFTVNDIYE 352
DB 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVEGFTVNDIYE 360
QY 353 FSPDQGRSKESGWVENEIYGY 374
DB 361 FSPDQGRSKESGWVENEIYGY 382

RESULT 11

US-09-904-011-137
; Sequence 137, Application US/09904011
; Publication No. US20030003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc

APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/904,011
CURRENT FILING DATE: 2001-07-11
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 137
LENGTH: 382
TYPE: PRT
ORGANISM: Homo Sapien
US-09-904-011-137

Query Match 99.3%; Score 1986; DB 11; Length 382;
Best Local Similarity 97.9%; Pred. No. 2.7e-178;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGGTORPCYKVIYFHDTSR 52
Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLDRGGQVPCVCRGGTQRCYKVIYFHDTSR 60
QY 53 RLNFEEAKEACRRDGGQLVSIESTSEDEQKLEKFTIENLLPSDGFWMGLRRREEKQSNSTA 112
Db 61 RLNFEEAKEACRRDGGQLVSIESTSEDEQKLEKFTIENLLPSDGFWMGLRRREEKQSNSTA 120
QY 113 QDLYAWTDGSIQFRNMYVDPSGSEVCVVMTHQPSAPAGIGGPFYMFQWDDRCNMKN 172
Db 121 QDLYAWTDGSIQFRNMYVDPSGSEVCVVMTHQPSAPAGIGGPFYMFQWDDRCNMKN 180
QY 173 NFICKYSDEKPAVPSREAGEETELTPVLPEETQBEDAKTKFKESREAAALNLAIVILIPS 232
Db 181 NFICKYSDEKPAVPSREAGEETELTPVLPEETQBEDAKTKFKESREAAALNLAIVILIPS 240
QY 233 IPLLALLVVTTCVWVICRKRREQDPSTKQHTIWPSPHQGNSPDLEVVYVIRKQSE 292
Db 241 IPLLALLVVTTCVWVICRKRREQDPSTKQHTIWPSPHQGNSPDLEVVYVIRKQSE 300
QY 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMNVNPPSSGGFVTLVSVESGFTNDIYE 352
Db 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMNVNPPSSGGFVTLVSVESGFTNDIYE 360
QY 353 FSPDQMGRSKESGWVENEIYGY 374
Db 361 FSPDQMGRSKESGWVENEIYGY 382

RESULT 12

US-09-906-742-137
Sequence 137, Application US/09906742
Publication No. US20030023054A1
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Ashkenazi, Avi
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/906,742
CURRENT FILING DATE: 2001-07-16
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222

Best Local Similarity 97.9%; Pred. No. 2.7e-178;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
Qy 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGTQPCYKVIYFHDTSR 52
Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSR 60
Qy 53 RLNFEEAKEACRRDGGQVLSIESDEQKLEKFTENLLPSDGDGFWIGLRREKQSNSTA 112
Db 61 RLNFEEAKEACRRDGGQVLSIESDEQKLEKFTENLLPSDGDGFWIGLRREKQSNSTA 120
Qy 113 QDLYAWTDGSIQFRNMYVDEPSCGSEVCMVYHQPSPAGIGGPTMFMQNDRCNMKN 172
Db 121 QDLYAWTDGSIQFRNMYVDEPSCGSEVCMVYHQPSPAGIGGPTMFMQNDRCNMKN 180
Qy 173 NFICKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNLAYILIPS 232
Db 181 NFICKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNLAYILIPS 240
Qy 233 IPLLALLVVTTCVWVWICRKRREQDPSTKKQHTIWPSPHQCNSPDLEVYVIRKQSE 292
Db 241 IPLLALLVVTTCVWVWICRKRREQDPSTKKQHTIWPSPHQCNSPDLEVYVIRKQSE 300
Qy 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYE 352
Db 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYE 360
Qy 353 FSPDQGRSKESGWENEIYGY 374
Db 361 FSPDQGRSKESGWENEIYGY 382

RESULT 14

US-09-907-613-137
; Sequence 137, Application US/09907613
; Publication No. US20030027145A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Deenoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertschen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,613
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26

; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-613-137

Query Match 99.3%; Score 1986; DB 11; Length 382;
Best Local Similarity 97.9%; Pred. No. 2.7e-178;
Matches 374; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Qy 1 MRPGTALQAVLLAVLLVGLRAATGRLLS-----GQVCRGGTQPCYKVIYFHDTSR 52
Db 1 MRPGTALQAVLLAVLLVGLRAATGRLLSASDLDRGGQPCVCRGGTQPCYKVIYFHDTSR 60
Qy 53 RLNFEEAKEACRRDGGQVLSIESDEQKLEKFTENLLPSDGDGFWIGLRREKQSNSTA 112
Db 61 RLNFEEAKEACRRDGGQVLSIESDEQKLEKFTENLLPSDGDGFWIGLRREKQSNSTA 120
Qy 113 QDLYAWTDGSIQFRNMYVDEPSCGSEVCMVYHQPSPAGIGGPTMFMQNDRCNMKN 172
Db 121 QDLYAWTDGSIQFRNMYVDEPSCGSEVCMVYHQPSPAGIGGPTMFMQNDRCNMKN 180
Qy 173 NFICKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNLAYILIPS 232
Db 181 NFICKYSDEKPAVPSREAGEETELTTPVLPETQEDAKKTFKESREAAALNLAYILIPS 240
Qy 233 IPLLALLVVTTCVWVWICRKRREQDPSTKKQHTIWPSPHQCNSPDLEVYVIRKQSE 292
Db 241 IPLLALLVVTTCVWVWICRKRREQDPSTKKQHTIWPSPHQCNSPDLEVYVIRKQSE 300
Qy 293 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYE 352
Db 301 ADLAETRPDLKNISFRVCSGEATPDDMSCDYDNMAVNPSESGFVTLVSVESGFVTNDIYE 360
Qy 353 FSPDQGRSKESGWENEIYGY 374
Db 361 FSPDQGRSKESGWENEIYGY 382

RESULT 15

US-09-907-942-137
; Sequence 137, Application US/09907942
; Publication No. US20030027146A1
; GENERAL INFORMATION:

CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 4 link domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U65989; AAC06238.2; -.
CC EMBL; S74662; AAC60527.1; -.
CC EMBL; L07054; -. NOT_ANNOTATED_CDS.
CC PIR; I46998; I46998.
CC HSP; P08709; IBF9.
CC InterPro; IPR00152; Asx hydroxyl.
CC InterPro; IPR00742; EGF_2.
CC InterPro; IPR01881; EGF_CA.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR01304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR003324; SGXSG.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 1.
CC Pfam; PF00047; Ig; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF02339; SGXSG; 66.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF0193; Xlink; 4.
CC PRINTS; PR01265; LINKMODULE.
CC PRINTS; PR00356; ANTIFREEZEII.
CC ProDom; PD000918; Link; 4.
CC SMART; SM00032; CCP; 1.
CC SMART; SM00034; CLECT; 1.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00445; LINK; 4.
CC PROSITE; PS00010; ASX-HYDROXYL; 1.
CC PROSITE; PS00615; C-TYPE LECTIN 1; 1.
CC PROSITE; PS00041; C-TYPE LECTIN 2; 1.
CC PROSITE; PS00022; EGF_1; 1.
CC PROSITE; PS01187; EGF_CA; 1.
CC PROSITE; PS00835; IG-LIKE; 1.
CC PROSITE; PS00290; IG-MHC; 1.
CC PROSITE; PS01241; LINK; 4.
CC Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
CC EGF-like domain; Repeat; Immunoglobulin domain.
CC SIGNAL 1 16 POTENTIAL.
CC CHAIN 17 2333 AGGRECAN CORE PROTEIN.
CC DOMAIN 34 147 IG-LIKE V-TYPE.
CC DOMAIN 170 247 LINK 1.
CC DOMAIN 268 349 LINK 2.
CC DOMAIN 513 590 LINK 3.
CC DOMAIN 611 692 LINK 4.
CC DOMAIN 2081 2117 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).
CC DOMAIN 2130 2245 C-TYPE LECTIN.
CC DOMAIN 2249 2307 SUSHI.
CC DOMAIN 48 140 GI-A.
CC DOMAIN 152 247 GI-B.
CC DOMAIN 253 349 GI-B'.
CC DOMAIN 495 589 G2-B.
CC DOMAIN 596 691 G2-B'.
CC DOMAIN 694 816 KS.

FT DOMAIN 819 1394 CS-1.
FT DOMAIN 1395 2079 CS-2.
FT DOMAIN 2080 2333 G3.
FT DISULFID 51 133 BY SIMILARITY.
FT DISULFID 175 246 BY SIMILARITY.
FT DISULFID 199 220 BY SIMILARITY.
FT DISULFID 273 348 BY SIMILARITY.
FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 518 589 BY SIMILARITY.
FT DISULFID 542 563 BY SIMILARITY.
FT DISULFID 616 691 BY SIMILARITY.
FT DISULFID 640 661 BY SIMILARITY.
FT DISULFID 2085 2096 BY SIMILARITY.
FT DISULFID 2090 2105 BY SIMILARITY.
FT DISULFID 2107 2116 BY SIMILARITY.
FT DISULFID 2123 2134 BY SIMILARITY.
FT DISULFID 2151 2243 BY SIMILARITY.
FT DISULFID 2219 2235 BY SIMILARITY.
FT DISULFID 2250 2293 BY SIMILARITY.
FT DISULFID 2279 2306 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 676 676 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2333 AA; 240573 MW; 8B9ED78F3508B596 CRC64;
Query Match 13.9%; Score 155.5; DB 1; Length 2333;
Best Local Similarity 28.5%; Pred. No. 1.4e-05;
Matches 51; Conservative 24; Mismatches 61; Indels 43; Gaps 11;
QY 9 QPVCRGG---TORPCYKVIYFHTDTSRRRLNFEAKACRRDGGQLVLSIESEDEQKLEKFI 65
Db 2120 QELCEEGWTKFGHCYR--YFPD---RESWVDAESRCRAQQSHLSIVTPEEQ----EFV 2170
QY 66 ENLLPSDGDGF-WIGLRRREKQSNSTACODLVAMTDGSIQSFRNMYVDEPS---CGSEVC 121
Db 2171 NN---NAQDYQWIGL-----NDRITGDFRSDGHSLSQFNWRPNQPDNFFVSGEDC 2219
QY 122 VVM-VHQPASAPAGIGGYPYFWMNDRCNMKNFKYKYS----DEKPAVPSREAGEETE 175
Db 2220 VVMIMHEKG-----ENMDVPCNYLPFTCKKGTVACGDPVVEHARTFGOKD 2267

Search completed: December 22, 2003, 16:10:59
Job time : 6.56161 secs